(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 October 2000 (05.10.2000)

PCT

(10) International Publication Number
WO 00/58473 A3

(51) International Patent Classification[7]: C12N 15/12, C07K 14/47, 16/18, G01N 33/566, C12Q 1/68, C12N 15/11, 15/62, A01K 67/027, A61K 38/00

(21) International Application Number: PCT/US00/08621

(22) International Filing Date: 31 March 2000 (31.03.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/127,607	31 March 1999 (31.03.1999)	US
60/127,636	2 April 1999 (02.04.1999)	US
60/127,728	5 April 1999 (05.04.1999)	US
09/540,763	30 March 2000 (30.03.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/127,607 (CIP)
Filed on	31 March 1999 (31.03.1999)
US	60/127,636 (CIP)
Filed on	2 April 1999 (02.04.1999)
US	60/127,728 (CIP)
Filed on	5 April 1999 (05.04.1999)
US	09/540,763 (CIP)
Filed on	30 March 2000 (30.03.2000)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SHIMKETS, Richard, A.** [US/US]; 191 Leete Street, West Haven, CT 06516 (US). **LEACH, Martin** [GB/US]; 884 School Street, Webster, MA 01570 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *With international search report.*

— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

(88) Date of publication of the international search report:
25 January 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/58473 A3

(54) Title: NUCLEIC ACIDS INCLUDING OPEN READING FRAMES ENCODING POLYPEPTIDES; "ORFX"

(57) Abstract: The present invention provides open reading frames encoding isolated polypeptides, as well as polynucleotides encoding ORFX and antibodies that immunospecifically bind to ORFX or any derivative, variant, mutant, or fragment of the ORFX polypeptides, polynucleotides or antibodies. The invention additionally provides methods in which the ORFX polypeptide, polynucleotide and antibody are used in detection and treatment of a broad range of pathological states, as well as to other uses.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/08621

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/47 C07K16/18 G01N33/566 C12Q1/68
C12N15/11 C12N15/62 A01K67/027 A61K38/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K G01N A01K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
BIOSIS, EMBASE, MEDLINE, CAB Data, PAJ, EPO-Internal, WPI Data, STRAND

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	COLE S.T.: "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence." NATURE, vol. 393, 11 June 1998 (1998-06-11), XP002144873 sequence ---	
A	LAMERDIN J.E.: "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine protease gene cluster." EMEST DATABASE ENTRY, 8 February 1999 (1999-02-08), XP002144874 sequence ---	

-/--

[X] Further documents are listed in the continuation of box C.

[ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 August 2000	23.11.00

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hix, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Internat' ! Application No
PCT/US 00/08621

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	M.D. ADAMS ET AL.: "The genome sequence of Drosophila melanogaster." SCIENCE, vol. 287, 24 March 2000 (2000-03-24), pages 2185-2195, XP002144875 the whole document -----	6

INTERNATIONAL SEARCH REPORT

Inte onal application No.
PCT/US 00/08621

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 27 to 32 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

claims 1 to 32 partially

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claim : 1 to 32 partially

Isolated nucleic acid molecule encoding a polypeptide comprising an amino acid sequence that is at least 85% identical to a polypeptide including an amino acid sequence selected from a group consisting of SEQ ID NO 2n wherein n is 1, oligonucleotides less than 100 nucleotides in length and comprising at least 6 contiguous nucleotides from the above sequence, polypeptides encoded by said nucleotides, antibodies that bind to said polypeptide, pharmaceutical composition comprising said polypeptide and methods of detection, screening, therapeutic uses involving said polypeptide.

2. Claim : .

Inventions 2 to 3161

claims 1 to 32 partially :

Isolated nucleic acid molecule encoding a polypeptide comprising an amino acid sequence that is at least 85% identical to a polypeptide including an amino acid sequence selected from a group consisting of SEQ ID NO 2n wherein n is 2 to 3161, oligonucleotides less than 100 nucleotides in length and comprising at least 6 contiguous nucleotides from the above sequence, polypeptides encoded by said nucleotides, antibodies that bind to said polypeptide, pharmaceutical composition comprising said polypeptide and methods of detection, screening, therapeutic uses involving said polypeptide.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY

BACKGROUND OF THE INVENTION

The invention relates generally to nucleic acids and polypeptides encoded thereby, and methods of using these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part on the discovery of nucleic acids that include open reading frames encoding novel polypeptides, and on the polypeptides encoded thereby. The nucleic acids and polypeptides are collectively referred to herein as "ORFX".

Accordingly, in one aspect, the invention provides an isolated nucleic acid molecule (SEQ ID NO:2*n*-1, wherein *n* is an integer between 1-3161), that encodes novel polypeptide, or a fragment, homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid sequence encoding a polypeptide at least 85% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO:2*n*, wherein *n* is an integer between 1-3161. The nucleic acid can be, *e.g.*, a genomic DNA fragment, or a cDNA molecule.

Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a recombinant expression vector comprising any of the nucleic acid molecules described above.

In another aspect, the invention includes a pharmaceutical composition that includes an ORFX nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified ORF polypeptide, *e.g.*, any of the ORFX polypeptides encoded by an ORFX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes a ORFX polypeptide and a pharmaceutically acceptable carrier or diluent.

In a still a further aspect, the invention provides an antibody that binds specifically to an ORFX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including ORFX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing an ORFX polypeptide by providing a cell containing a ORFX nucleic acid, *e.g.*, a vector that includes a ORFX nucleic acid, and culturing the cell under conditions sufficient to express the ORFX polypeptide encoded by the nucleic acid. The expressed ORFX polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous ORFX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying an ORFX polypeptide or nucleic acids in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of a ORFX polypeptide by contacting ORFX polypeptide with a compound and determining whether the ORFX polypeptide activity is modified.

The invention is also directed to compounds that modulate ORFX polypeptide activity identified by contacting a ORFX polypeptide with the compound and determining whether the compound modifies activity of the ORFX polypeptide, binds to the ORFX polypeptide, or binds to a nucleic acid molecule encoding a ORFX polypeptide.

In a another aspect, the invention provides a method of determining the presence of or predisposition of an ORFX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of ORFX polypeptide in the subject sample.

The amount of ORFX polypeptide in the subject sample is then compared to the amount of ORFX polypeptide in a control sample. An alteration in the amount of ORFX polypeptide in the subject protein sample relative to the amount of ORFX polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition. Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the ORFX is detected using a ORFX antibody.

In a further aspect, the invention provides a method of determining the presence of or predisposition of an ORFX-associated disorder in a subject. The method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the ORFX nucleic acid in the subject nucleic acid sample. The amount of ORFX nucleic acid sample in the subject nucleic acid is then compared to the amount of an ORFX nucleic acid in a control sample. An alteration in the amount of ORFX nucleic acid in the sample relative to the amount of ORFX in the control sample indicates the subject has a tissue proliferation-associated disorder.

In a still further aspect, the invention provides method of treating or preventing or delaying a ORFX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired a ORFX nucleic acid, a ORFX polypeptide, or an ORFX antibody in an amount sufficient to treat, prevent, or delay a tissue proliferation-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides novel polypeptides and nucleotides encoded thereby. The polynucleotides and their encoded polypeptides can be grouped according to the functions played by their gene products. Such functions include, structural proteins, proteins from which associated with metabolic pathways fatty acid metabolism, glycolysis, intermediary metabolism, calcium metabolism, proteases, and amino acid metabolism, etc.

Included in the invention are 3161 novel nucleic acid sequences and their encoded polypeptides. The sequences are collectively referred to as "ORFX nucleic acids" or ORFX polynucleotides" and the corresponding encoded polypeptide is referred to as a "ORFX polypeptide" or ORFX protein". For example, an ORFX nucleic acid according to the invention is a nucleic acid including an ORF1 nucleic acid, and an ORF polypeptide according to the invention is a polypeptide that includes the amino acid sequence of an ORF1 polypeptide. Unless indicated otherwise, "ORFX" is meant to refer to any of the ORF1-3161 sequences disclosed herein.

Table 1 provides a summary of the ORFX nucleic acids and their encoded polypeptides are summarized in Table 1. Nucleic acid sequences and polypeptide sequences for ORFX nucleic acids according to the invention is provided in the section of the specification entitled "Disclosed Sequences of ORFX Nucleic Acid and Polypeptide Sequences."

Column 1 of Table 1, entitled "ORF #", denotes an ORF number assigned to a nucleic acid containing an open reading frame according to the invention.

Column 2 of Table 1, entitled "Internal Identification number (Nucleic Acid Sequence Identification Number, Polypeptide Sequence Identification Number), provides an internal identification number for the indicated ORF, along with sequence identification numbers (SEQ ID NOs.) corresponding to the indicated ORF. In general, for an ORF*n* according to the invention (wherein *n* is any integer from 1 to 3161), a nucleic acid corresponding to the ORF is SEQ ID NO:2*n*-1, and an amino acid sequence encoded by the ORF is SEQ ID NO:2*n*. For example, a nucleic acid sequence corresponding to an ORF1 nucleic acid is SEQ ID NO:1, and a polypeptide sequence corresponding to an ORF1 polypeptide is SEQ ID NO:2. Similarly, a

nucleic acid sequence corresponding to an ORF4 nucleic acid is SEQ ID NO:7, and a polypeptide sequence corresponding to an ORF4 polypeptide is SEQ ID NO:8; a nucleic acid sequence corresponding to an ORF198 nucleic acid sequence is SEQ ID NO:395, and a polypeptide sequence corresponding to an ORF198 polypeptide is SEQ ID NO:396. Nucleic acid sequences and polypeptide sequences for ORFX nucleic acids according to the invention are provided in the section of the specification entitled "Disclosed Sequences of ORFX Nucleic Acid and Polypeptide Sequences."

Column 2 of Table 1, entitled "Protein Similarity", lists previously described proteins that are related to polypeptides encoded by the ORFs. Genbank identifiers for the previously described proteins are provided. These can be retrieved from http://www.ncbi.nlm.nih.gov/.

To determine similarity to previously described proteins, polypeptides encoded by ORFX DNA sequences were tested using the Framesearch Algorithm against a nonredundant version of the GenPept Database from NCBI/Genbank. DNA sequences that had a score of '90' or above (Framesearch algorithm score, Edelman et. al. GCG Genetics) to a known protein were selected. Open reading frames were extended beyond the region of the protein matched using standard DNA translation and codon tables. Novel proteins that lacked a protein match were translated against the standard genetic codons and proteins with an ORF at least 80 amino acids and containing a Methionine start are included in the Table.

Column 3 of Table 3, entitled "Protein Domains", lists previously described protein domains, designated by pfam entries, that are present in polypeptides encoded by the ORFs, Also included in column 3 are proteins in which these domains are present. The pfam entries can be retrieved from http://pfam.wustl.edu/. DNA sequences were translated in all six frames and tested using the Hmmer Algorithm against the Pfam Database (References to the algorithm and Pfam database can be found at http://pfam.wustl.edu). Translated DNA sequences that matched a protein domain entry in the Pfam database AND had a score of'7.5' were selected.

Column 4 of Table 3, entitled "Protein Classification", lists the type of classification assigned for the protein, based on its homology. Examples of proteins in the classification include the following proteins:

Amylases

Amylase is responsible for endohydrolysis of 1,4-alpha-glucosidic linkages in oligosaccharides and polysaccharides. Variations in amylase gene may be indicative of delayed maturation and of various amylase producing neoplasms and carcinomas.

Amyloid

The serum amyloid A (SAA) proteins comprise a family of vertebrate proteins that associate predominantly with high density lipoproteins (HDL). The synthesis of certain members of the family is greatly increased in inflammation. Prolonged elevation of plasma SAA levels, as in chronic inflammation, 15 results in a pathological condition, called amyloidosis, which affects the liver, kidney and spleen and which is characterized by the highly insoluble accumulation of SAA in these tissues. Amyloid selectively inhibits insulin-stimulated glucose utilization and glycogen deposition in muscle, while not affecting adipocyte glucose metabolism. Deposition of fibrillar amyloid proteins intraneuronally, as neurofibrillary tangles, extracellularly, as plaques and in blood vessels, is characteristic of both Alzheimer's disease and aged Down's syndrome. Amyloid deposition is also associated with type II diabetes mellitus.

Angiopoeitin

Members of the angiopoeitin/fibrinogen family have been shown to stimulate the generation of new blood vessels, inhibit the generation of new blood vessels, and perform several roles in blood clotting. This generation of new blood vessels, called angiogenesis, is also an essential step in tumor growth in order for the tumor to get the blood supply it needs to expand. Variation in these genes may be predictive of any form of heart disease, numerous blood clotting disorders, stroke, hypertension and predisposition to tumor formation and metastasis. In particular, these variants may be predictive of the response to various antihypertensive drugs and chemotherapeutic and anti-tumor agents.

Apoptosis-related proteins

Active cell suicide (apoptosis) is induced by events such as growth factor withdrawal and toxins. It is controlled by regulators, which have either an inhibitory effect on programmed cell

death (anti-apoptotic) or block the protective effect of inhibitors (pro-apoptotic). Many viruses have found a way of countering defensive apoptosis by encoding their own anti-apoptosis genes preventing their target-cells from dying too soon. Variants of apoptosis related genes may be useful in formulation of anti-aging drugs.

Cadherin, Cyclin, Polymerase, Oncogenes, Histones, Kinases

Members of the cell division/cell cycle pathways such as cyclins, many transcription factors and kinases, DNA polymerases, histones, helicases and other oncogenes play a critical role in carcinogenesis where the uncontrolled proliferation of cells leads to tumor formation and eventually metastasis. Variation in these genes may be predictive of predisposition to any form of cancer, from increased risk of tumor formation to increased rate of metastasis. In particular, these variants may be predictive of the response to various chemotherapeutic and anti-tumor agents.

Colony-stimulating factor-related proteins

Granulocyte/macrophage colony-stimulating factors are cytokines that act in hematopoiesis by controlling the production, differentiation, and function of 2 related white cell populations of the blood, the granulocytes and the monocytes-macrophages.

Complement-related proteins

Complement proteins are immune associated cytotoxic agents, acting in a chain reaction to exterminate target cells to that were opsonized (primed) with antibodies, by forming a membrane attack complex (MAC). The mechanism of killing is by opening pores in the target cell membrane. Variations in 20 complement genes or their inhibitors are associated with many autoimmune disorders. Modified serum levels of complement products cause edemas of various tissues, lupus (SLE), vasculitis, glomerulonephritis, renal failure, hemolytic anemia, thrombocytopenia, and arthritis. They interfere with mechanisms of ADCC (antibody dependent cell cytotoxicity), severely impair immune competence and reduce phagocytic ability. Variants of complement genes may also be indicative of type I diabetes mellitus, meningitis neurological disorders such as nemaline myopathy, neonatal hypotonia, muscular disorders such as congenital myopathy and other diseases.

Cytochrome

The respiratory chain is a key biochemical pathway which is essential to all aerobic cells. There are five different cytochromes involved in the chain. These are heme bound proteins which serve as electron carriers. Modifications in these genes may be predictive of ataxia areflexia, dementia and myopathic and neuropathic changes in muscles. Also, association with various types of solid tumors.

Kinesins

Kinesins are tubulin molecular motors that function to transport organelles within cells and to move chromosomes along microtubules during cell division. Modifications of these genes may be indicative of neurological disorders such as Pick disease of the brain, tuberous sclerosis.

Cytokines, Interferon, Interleukin

Members of the cytokine families are known for their potent ability to stimulate cell growth and division even at low concentrations. Cytokines such as erythropoietin are cell-specific in their growth stimulation; erythropoietin is useful for the stimulation of the proliferation of erythroblasts. Variants in cytokines may be predictive for a wide variety of diseases, including cancer predisposition.

G-protein coupled receptors

G-protein coupled receptors (also called R7G) are an extensive group of hormones, neurotransmitters, odorants and light receptors which transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins. Alterations in genes coding for G-coupled proteins may be involved in and indicative of a vast number of physiological conditions. These include blood pressure regulation, renal dysfunctions, male infertility, dopamine associated cognitive, emotional, and endocrine functions, hypercalcemia, chondrodysplasia and osteoporosis, pseudohypoparathyroidism, growth retardation and dwarfism.

Thioesterases

Eukaryotic thiol proteases are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. Variants of thioester associated genes may be predictive of neuronal disorders and mental illnesses such as Ceroid Lipoffiscinosis, Neuronal 1, Infantile, Santavuori disease and more.

The key to the molecule type is as follows:

Abbrev:	Title:
amylase	amylase protein
amylaseinhib	amylase inhibitor
amyloid	amyloid protein
apoptosis	apoptosis associated protein
apoptosisinhib	apoptosis inhibitors
apoptosisrecep	apoptosis receptors
ATPase_associated	ATPase associated protein
biotindep	biotin dependent enzyme/protein
cadherin	cadherin protein
calcium_channel	calcium channel protein
carboxylase	carboxylase protein
cathepsin	cathepsin/carboxypeptidases
cathepsininhib	cathepsin/carboxypeptidase inhibitor
chloride_channel	chloride channel protein
collagen	collagen
complement	complement protein
complementrecept	complement receptor protein
complementinhib	complement inhibitor
csf	colony stimulating factor
csfrecept	colony stimulating factor receptor
cyclin	cyclin protein
cyto450	cytochrome p450 protein
cytochrome	cytochrome related protein
deaminase	deaminase
dehydrogenase	dehydrogenase
desaturase	desaturase
dna_rna_bind	DNA/RNA binding protein/factor
dna_rna_inhib	DNA/RNA binding protein/factor inhibitor
dynein	dynein

elastase	elastase
elastaseinhib	elastase inhibitor
eph	EPH family of tyrosine kinases
esterase	esterase
esteraseinhib	esterase inhibitor
fgf	fibroblast growth factor
fgfreceptor	fibroblast growth factor receptor
gaba	GABA receptor
glucoamylase	glucoamylase
glucoronidase	glucoronidase
glycoprotein	glycoprotein
Guanylyl	guanylylate cyclase
helicase	helicase
histone	histone
HOM	homologous
homeobox	homeobox protein
hydrolase	hydrolase
hydroxysteroid	hydroxysteroid associated protein
hypoxanthine	hypoxanthine associated protein
immunoglob	immunoglobulin
immunoglobrecept	immunoglobulin receptor
interferon	interferon
interleukin	interleukin
interleukinrecept	interleukin receptor
isomerase	isomerase
isomeraseinhibitor	isomerase inhibitor
isomerasereceptor	isomerase receptor
kinase	kinase
kinaseinhibitor	kinase inhibitor
kinasereceptor	kinase receptor
kinesin	kinesin
laminin	laminin associated protein
lipase	lipase
metallothionein	metallothionein
MHC	major histocompatability complex
misc_channel	miscellaneous channel
ngf	nerve growth factor
nuci_recpt	nuclear receptor
nuclease	nuclease
oncogene	oncogene associated protein
oxidase	oxidase
oxygenase	oxygenase
peptidase	peptidase
peroxidase	peroxidase
phosphatase	phosphatase
phosphataseinhib	phosphatase inhibitor

phosphorylase	phosphorylase
PIR	PIR DATABASE (release 56, 29-OCT-1998)
polymerase	polymerase
potassium_channel	potassium channel protein
prostaglandin	prostaglandin
protease	protease
proteaseinhib	protease inhibitor
reductase	reductase
ribosomalprot	ribosomal associated protein
RTR	EMBLDATABASE translated entries not to be incorporated into SWISS-PROT (20-JUL-1998)
SIM	similar
SPTR	EMBL DATABASE translated entries to be incorporated into SWISS-PROT (20-JUL-1998)
struct	structural associated protein
sulfotransferase	sulfotransferase
SWP	SWISS-PROT DATABASE (release 18-OCT-1998)
SWPN	SWISS-PROT Update (release 11-NOV-98)
synthase	synthase
tgf	transforming growth factor
tgfreceptor	transforming growth factor receptor
thioesterase	thioesterase
thiolase	thiolase
tm7	seven transmembrane domain G-protein coupled receptor
tnf	necrosis factor receptor
traffic	tumor necrosis factor
tnfreceptor	tumor trafficking associated protein
TRN	EMBL DATABASE translated entries update (20-JUL-1998)
transcriptfactor	transcription factor
transferase	transferase
transport	transport protein
tubulin	tubulin
ubiquitin	ubiquitin
unclassified	Protein not categorized into one of the aforementioned protein families
water channel	water channel protein

Column 5 of Table 1, entitled, "Cells or Tissues in Which Gene is Expressed", denotes tissues, represented by five digit numbers, in which RNA homologous to the ORF nucleic acid sequences is present. Tissues or cells corresponding to the numbers are provided in Table 2.

ORFX nucleic acids, and their encoded polypeptides, according to the invention are useful in a variety of applications and contexts. For example, various ORFX nucleic acids and polypeptides according to the invention are useful, *inter alia*, as novel members of the protein families indicated in Table 1, and/or according to the presence of domains and sequence relatedness to previously described proteins as summarized in Table 1.

ORFX nucleic acids and polypeptides according to the invention can also be used to identify cell types listed in Table 1 for an indicated ORFX according to the invention. Additional utilities for ORFX nucleic acids and polypeptides according to the invention are disclosed herein.

ORFX Nucleic Acids

The novel nucleic acids of the invention include those that encode an ORFX or ORFX-like protein, or biologically active portions thereof. The nucleic acids include nucleic acids encoding polypeptides that include the amino acid sequence of one or more of SEQ ID NO:2*n*, wherein *n* = 1 to 3161. The encoded polypeptides can thus include, *e.g.*, the amino acid sequences of SEQ ID NO: 2, 4, 6, 8, 10, . . ., 6310, 6312, 6314, 6316, 6318, 6320, and/or 6322.

In some embodiments, a nucleic acid encoding a polypeptide having the amino acid sequence of one or more of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) includes the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a fragment thereof. Additionally, the invention includes mutant or variant nucleic acids of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a fragment thereof, any of whose bases may be changed from the disclosed sequence while still encoding a protein that maintains its ORFX -like activities and physiological functions. The invention further includes the complement of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), including fragments, derivatives,

analogs and homolog thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

Also included are nucleic acid fragments sufficient for use as hybridization probes to identify ORFX-encoding nucleic acids (*e.g.*, ORFX mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of ORFX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated ORFX nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a complement of any of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n*=1 to 3161) as a hybridization probe, ORFX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to ORFX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at lease 6 contiguous nucleotides of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a complement thereof. Oligonucleotides may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in

is one that is sufficiently complementary to the nucleotide sequence shown in of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161),, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of ORFX. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the

invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489, which is incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of ORFX polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a ORFX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human ORFX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161) as well as a polypeptide having ORFX activity. Biological activities of the ORFX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human ORFX polypeptide.

The nucleotide sequence determined from the cloning of the human ORFX gene allows for the generation of probes and primers designed for use in identifying the cell types disclosed and/or cloning ORFX homologues in other cell types, *e.g.*, from other tissues, as well as ORFX homologues from other mammals. The probe/primer typically comprises a substantially purified

oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161); or an anti-sense strand nucleotide sequence of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161); or of a naturally occurring mutant of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161).

Probes based on the human ORFX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a ORFX protein, such as by measuring a level of a ORFX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting ORFX mRNA levels or determining whether a genomic ORFX gene has been mutated or deleted.

"A polypeptide having a biologically active portion of ORFX" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of ORFX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), that encodes a polypeptide having a ORFX biological activity (biological activities of the ORFX proteins are summarized in Table 1), expressing the encoded portion of ORFX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of ORFX. For example, a nucleic acid fragment encoding a biologically active portion of ORFX can optionally include a domain as shown in Table 1, column 4.

ORFX variants

The invention further encompasses nucleic acid molecules that differ from the disclosed ORFX nucleotide sequences due to degeneracy of the genetic code. These nucleic acids thus encode the same ORFX protein as that encoded by the nucleotide sequence shown in SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161). In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in any of SEQ ID NO:2*n* (wherein *n* = 1 to 3161).

In addition to the human ORFX nucleotide sequence shown in any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of ORFX may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the ORFX gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a ORFX protein, preferably a mammalian ORFX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the ORFX gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in ORFX that are the result of natural allelic variation and that do not alter the functional activity of ORFX are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding ORFX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the ORFX cDNAs of the invention can be isolated based on their homology to the human ORFX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

In another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein n = 1 to 3161). In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding ORFX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are

hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981, *Proc Natl Acad Sci USA 78*: 6789-6792.

Conservative mutations

In addition to naturally-occurring allelic variants of the ORFX sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of any of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), thereby leading to changes in the amino acid sequence of the encoded ORFX protein, without altering the functional ability of the ORFX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of any of SEQ ID NO:2n-1 (wherein *n* = 1 to 3161). A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of ORFX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the ORFX proteins of the present invention, are predicted to be particularly unamenable to alteration.

Amino acid residues that are conserved among members of an ORFX family members are predicted to be less amenable to alteration. For example, an ORFX protein according to the present invention can contain at least one domain (*e.g.*, as shown in Table 1) that is a typically conserved region in an ORFX family member. As such, these conserved domains are not likely to be amenable to mutation. Other amino acid residues, however, (*e.g.*, those that are not conserved or only semi-conserved among members of the ORFX family) may not be as essential for activity and thus are more likely to be amenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding ORFX proteins that contain changes in amino acid residues that are not essential for activity. Such ORFX proteins differ in amino acid sequence from any of any of SEQ ID NO:2*n* (wherein n = 1 to 3161), yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of any of SEQ ID NO:2*n* (wherein n = 1 to 3161). Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to any of SEQ ID NO:2*n* (wherein n = 1 to 3161), more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO:2.

An isolated nucleic acid molecule encoding a ORFX protein homologous to the protein of any of SEQ ID NO:2*n* (wherein n = 1 to 3161) can be created by introducing one or more nucleotide substitutions, additions or deletions into the corresponding nucleotide sequence, *i.e.* SEQ ID NO:2*n*-1 for the corresponding *n*, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NO:2*n*-1 (wherein n = 1 to 3161) by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline,

phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in ORFX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a ORFX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for ORFX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant ORFX protein can be assayed for (1) the ability to form protein:protein interactions with other ORFX proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant ORFX protein and a ORFX receptor; (3) the ability of a mutant ORFX protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g.*, avidin proteins); (4) the ability to bind BRA protein; or (5) the ability to specifically bind an anti-ORFX protein antibody.

Antisense

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1 (wherein n = 1 to 3161), or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire ORFX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a ORFX protein of any of SEQ ID NO:2*n* (wherein n = 1 to 3161) or antisense nucleic acids complementary to a ORFX nucleic acid sequence of SEQ ID NO:2n-1 (wherein n = 1 to 3161) are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding ORFX. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid

residues (*e.g.*, the protein coding region of a human ORFX that corresponds to any of SEQ ID NO:2*n* (wherein n = 1 to 3161)). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding ORFX. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding ORFX disclosed herein (*e.g.*, SEQ ID NO:2n-1 (wherein n = 1 to 3161)), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of ORFX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of ORFX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of ORFX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil,

3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a ORFX protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

Ribozymes and PNA moieties

Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are

carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave ORFX mRNA transcripts to thereby inhibit translation of ORFX mRNA. A ribozyme having specificity for a ORFX-encoding nucleic acid can be designed based upon the nucleotide sequence of a ORFX DNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161)). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a ORFX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, ORFX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, ORFX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the ORFX (*e.g.*, the ORFX promoter and/or enhancers) to form triple helical structures that prevent transcription of the ORFX gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of ORFX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide

synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of ORFX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of ORFX can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of ORFX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of ORFX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, Proc. *Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or

the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

ORFX polypeptides

The novel protein of the invention includes the ORFX-like protein whose sequence is provided in any of SEQ ID NO:2*n* (wherein n = 1 to 3161). The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in FIG. 1 while still encoding a protein that maintains its ORFX-like activities and physiological functions, or a functional fragment thereof. For example, the invention includes the polypeptides encoded by the variant ORFX nucleic acids described above. In the mutant or variant protein, up to 20% or more of the residues may be so changed.

In general, an ORFX -like variant that preserves ORFX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above. Furthermore, without limiting the scope of the invention, positions of any of SEQ ID NO:2*n* (wherein n = 1 to 3161) may be substitute such that a mutant or variant protein may include one or more substitutions

The invention also includes isolated ORFX proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-ORFX antibodies. In one embodiment, native ORFX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, ORFX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a ORFX

protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the ORFX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of ORFX protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of ORFX protein having less than about 30% (by dry weight) of non-ORFX protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-ORFX protein, still more preferably less than about 10% of non-ORFX protein, and most preferably less than about 5% non-ORFX protein. When the ORFX protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of ORFX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of ORFX protein having less than about 30% (by dry weight) of chemical precursors or non-ORFX chemicals, more preferably less than about 20% chemical precursors or non-ORFX chemicals, still more preferably less than about 10% chemical precursors or non-ORFX chemicals, and most preferably less than about 5% chemical precursors or non-ORFX chemicals.

Biologically active portions of a ORFX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the ORFX protein, *e.g.*, the amino acid sequence shown in SEQ ID NO:2 that include fewer amino acids than the full length ORFX proteins, and exhibit at least one activity of a ORFX protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the ORFX protein. A biologically active portion of a ORFX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a ORFX protein of the present invention may contain at least one of the above-identified domains conserved between the FGF family of proteins. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native ORFX protein.

In an embodiment, the ORFX protein has an amino acid sequence shown in any of SEQ ID NO:2*n* (wherein n = 1 to 3161). In other embodiments, the ORFX protein is substantially homologous to any of SEQ ID NO:2*n* (wherein n = 1 to 3161) and retains the functional activity of the protein of any of SEQ ID NO:2*n* (wherein n = 1 to 3161), yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below. Accordingly, in another embodiment, the ORFX protein is a protein that comprises an amino acid sequence at least about 45% homologous, and more preferably about 55, 65, 70, 75, 80, 85, 90, 95, 98 or even 99% homologous to the amino acid sequence of any of SEQ ID NO:2*n* (wherein n = 1 to 3161) and retains the functional activity of the ORFX proteins of the corresponding polypeptide having the sequence of SEQ ID NO:2*n* (wherein n = 1 to 3161).

Determining homology between two or more sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See, *Needleman and Wunsch* 1970 *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a

degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NO:2*n*-1 (wherein n = 1 to 3161).

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of positive residues.

Chimeric and fusion proteins

The invention also provides ORFX chimeric or fusion proteins. As used herein, a ORFX "chimeric protein" or "fusion protein" includes a ORFX polypeptide operatively linked to a non-ORFX polypeptide. A "ORFX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to ORFX, whereas a "non-ORFX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the ORFX protein, *e.g.*, a protein that is different from the ORFX protein and that is derived from the same or a different organism. Within a ORFX fusion protein the ORFX polypeptide can correspond to all or a portion of a ORFX protein. In one embodiment, a ORFX fusion protein comprises at least one biologically active portion of a ORFX protein. In another embodiment, a ORFX fusion protein comprises at least two biologically active portions of a

ORFX protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the ORFX polypeptide and the non-ORFX polypeptide are fused in-frame to each other. The non-ORFX polypeptide can be fused to the N-terminus or C-terminus of the ORFX polypeptide.

For example, in one embodiment a ORFX fusion protein comprises a ORFX polypeptide operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds that modulate ORFX activity (such assays are described in detail below).

In another embodiment, the fusion protein is a GST-ORFX fusion protein in which the ORFX sequences are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant ORFX.

In yet another embodiment, the fusion protein is a ORFX protein containing a heterologous signal sequence at its N-terminus. For example, the native ORFX signal sequence can be removed and replaced with a signal sequence from another protein. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of ORFX can be increased through use of a heterologous signal sequence.

In another embodiment, the fusion protein is a ORFX-immunoglobulin fusion protein in which the ORFX sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The ORFX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ORFX ligand and a ORFX protein on the surface of a cell, to thereby suppress ORFX-mediated signal transduction *in vivo*. In one nonlimiting example, a contemplated ORFX ligand of the invention is an ORFX receptor. The ORFX-immunoglobulin fusion proteins can be used to modulate the bioavailability of a ORFX cognate ligand. Inhibition of the ORFX ligand/ORFX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the ORFX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-ORFX antibodies in a subject, to purify ORFX ligands, and in screening assays to identify molecules that inhibit the interaction of ORFX with a ORFX ligand.

A ORFX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A ORFX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the ORFX protein.

ORFX agonists and antagonists

The present invention also pertains to variants of the ORFX proteins that function as either ORFX agonists (mimetics) or as ORFX antagonists. Variants of the ORFX protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the ORFX protein. An agonist of the ORFX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the ORFX protein. An antagonist of the ORFX protein can inhibit one or more of the activities of the naturally occurring form of the ORFX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the ORFX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the ORFX proteins.

Variants of the ORFX protein that function as either ORFX agonists (mimetics) or as ORFX antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the ORFX protein for ORFX protein agonist or antagonist activity. In one

embodiment, a variegated library of ORFX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of ORFX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential ORFX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of ORFX sequences therein. There are a variety of methods which can be used to produce libraries of potential ORFX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential ORFX sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477.

Polypeptide libraries

In addition, libraries of fragments of the ORFX protein coding sequence can be used to generate a variegated population of ORFX fragments for screening and subsequent selection of variants of a ORFX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a ORFX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the ORFX protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of ORFX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors,

transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify ORFX variants (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).

Anti-ORFX Antibodies

The invention further encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the proteins of the invention.

An isolated ORFX protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind ORFX using standard techniques for polyclonal and monoclonal antibody preparation. Full-length ORFX protein can be used. Alternatively, the invention provides antigenic peptide fragments of ORFX for use as immunogens. The antigenic peptide of ORFX comprises at least 4 amino acid residues of the amino acid sequence shown in any of SEQ ID NO:2*n* (wherein n = 1 to 3161). The antigenic peptide encompasses an epitope of ORFX such that an antibody raised against the peptide forms a specific immune complex with ORFX. The antigenic peptide may comprise at least 6 aa residues, at least 8 aa residues, at least 10 aa residues, at least 15 aa residues, at least 20 aa residues, or at least 30 aa residues. In one embodiment of the invention, the antigenic peptide comprises a polypeptide comprising at least 6 contiguous amino acids of any of SEQ ID NO:2*n* (wherein n = 1 to 3161).

In an embodiment of the invention, epitopes encompassed by the antigenic peptide are regions of ORFX that are located on the surface of the protein, *e.g.*, hydrophilic regions. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, Proc. Nat. Acad. Sci. USA 78: 3824-3828; Kyte and Doolittle 1982, J. Mol. Biol. 157: 105-142, each incorporated herein by reference in their entirety.

As disclosed herein, an ORFX protein sequence of any of SEQ ID NO:2*n* (wherein n = 1 to 3161) , or derivatives, fragments, analogs or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind these protein components. The term "antibody" as used herein refers to immunoglobulin molecules and

immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen, such as ORFX. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In a specific embodiment, antibodies to human ORFX proteins are disclosed. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to a ORFX protein sequence of any of SEQ ID NO:2*n* (wherein n = 1 to 3161) or derivative, fragment, analog or homolog thereof. Some of these proteins are discussed below.

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, recombinantly expressed ORFX protein or a chemically synthesized ORFX polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against ORFX can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of ORFX. A monoclonal antibody composition thus typically displays a single binding affinity for a particular ORFX protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular ORFX protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (see Kohler & Milstein, 1975 *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (see Kozbor, *et al.*, 1983 *Immunol Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND

CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, *et al.*, 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (see Cole, *et al.*, 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Each of the above citations are incorporated herein by reference in their entirety

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to a ORFX protein (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see *e.g.*, Huse, *et al.*, 1989 *Science* 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a ORFX protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. See *e.g.*, U.S. Patent No. 5,225,539. Each of the above citations are incorporated herein by reference. Antibody fragments that contain the idiotypes to a ORFX protein may be produced by techniques known in the art including, but not limited to: (*i*) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (*ii*) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (*iii*) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (*iv*) F_v fragments.

Additionally, recombinant anti-ORFX antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 86/01533; U.S. Pat. No. 4,816,567; European Patent Application No. 125,023; Better *et al.*(1988) *Science* 240:1041-1043; Liu *et al.* (1987) *PNAS* 84:3439-3443; Liu *et al.* (1987) *J Immunol.* 139:3521-3526; Sun *et al.* (1987) *PNAS* 84:214-218; Nishimura *et al.* (1987) *Cancer Res* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; Shaw *et al.* (1988), *J. Natl Cancer Inst* 80:1553-1559); Morrison(1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; U.S. Pat. No. 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525;

Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J Immunol* 141:4053-4060. Each of the above citations are incorporated herein by reference.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of a ORFX protein is facilitated by generation of hybridomas that bind to the fragment of a ORFX protein possessing such a domain. Antibodies that are specific for one or more domains within a ORFX protein, *e.g.*, the domain spanning the first fifty amino-terminal residues specific to ORFX when compared to FGF-9, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-ORFX antibodies may be used in methods known within the art relating to the localization and/or quantitation of a ORFX protein (*e.g.*, for use in measuring levels of the ORFX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for ORFX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody derived binding domain, are utilized as pharmacologically-active compounds [hereinafter "Therapeutics"].

An anti-ORFX antibody (*e.g.*, monoclonal antibody) can be used to isolate ORFX by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-ORFX antibody can facilitate the purification of natural ORFX from cells and of recombinantly produced ORFX expressed in host cells. Moreover, an anti-ORFX antibody can be used to detect ORFX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the ORFX protein. Anti-ORFX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or

phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

ORFX Recombinant Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding ORFX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements

(*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, ORFX proteins, mutant forms of ORFX, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of ORFX in prokaryotic or eukaryotic cells. For example, ORFX can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (1) to increase expression of recombinant protein; (2) to increase the solubility of the recombinant protein; and (3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the ORFX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari, *et al.*, (1987) *EMBO J* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, ORFX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith *et al.* (1983) *Mol Cell Biol* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells. See, *e.g.*, Chapters 16 and 17 of Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv Immunol* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to ORFX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub *et al.*, "Antisense RNA as a molecular tool for genetic analysis," Reviews--Trends in Genetics, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant

host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, ORFX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding ORFX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) ORFX protein. Accordingly, the invention further provides methods for producing ORFX protein using the host cells of the invention. In one embodiment,

the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding ORFX has been introduced) in a suitable medium such that ORFX protein is produced. In another embodiment, the method further comprises isolating ORFX from the medium or the host cell.

Transgenic animals

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which ORFX-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous ORFX sequences have been introduced into their genome or homologous recombinant animals in which endogenous ORFX sequences have been altered. Such animals are useful for studying the function and/or activity of ORFX and for identifying and/or evaluating modulators of ORFX activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous ORFX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing ORFX-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The human ORFX DNA sequence of SEQ ID NO:2*n*-1 (wherein $n = 1$ to 3161) can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human ORFX gene, such as a mouse ORFX gene, can be isolated based on hybridization to the human ORFX cDNA (described further above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of

expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the ORFX transgene to direct expression of ORFX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan 1986, In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the ORFX transgene in its genome and/or expression of ORFX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding ORFX can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a ORFX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the ORFX gene. The ORFX gene can be a human gene (*e.g.*, SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161)), but more preferably, is a non-human homologue of a human ORFX gene. For example, a mouse homologue of human ORFX gene of SEQ ID NO:2n-1 (wherein *n* = 1 to 3161) can be used to construct a homologous recombination vector suitable for altering an endogenous ORFX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous ORFX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous ORFX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous ORFX protein). In the homologous recombination vector, the altered portion of the ORFX gene is flanked at its 5' and 3' ends by additional nucleic acid of the ORFX gene to allow for homologous recombination to occur between the exogenous ORFX gene carried by the vector and an endogenous ORFX gene in an embryonic stem cell. The additional flanking ORFX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector. See *e.g.*, Thomas *et al.* (1987) *Cell* 51:503 for a description of

homologous recombination vectors. The vector is introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced ORFX gene has homologously recombined with the endogenous ORFX gene are selected (see *e.g.*, Li *et al.* (1992) *Cell* 69:915).

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. See *e.g.*, Bradley 1987, In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Curr Opin Biotechnol* 2:823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, *e.g.*, Lakso *et al.* (1992) *PNAS* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.* (1997) *Nature* 385:810-813. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of

this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

Pharmaceutical Compositions

The ORFX nucleic acid molecules, ORFX proteins, and anti-ORFX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a ORFX protein or anti-ORFX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use

as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by any of a number of routes, *e.g.*, as described in U.S. Patent Nos. 5,703,055. Delivery can thus also include, *e.g.*, intravenous injection, local administration (see U.S. Pat. No. 5,328,470) or stereotactic injection (see *e.g.*, Chen *et al.* (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Additional Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: (a) screening assays; (b) detection assays (*e.g.*, chromosomal mapping, cell and tissue typing, forensic biology), (c) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and (d) methods of treatment (*e.g.*, therapeutic and prophylactic).

The isolated nucleic acid molecules of the invention can be used to express ORFX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect ORFX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a ORFX gene, and to modulate ORFX activity, as described further below. In addition, the ORFX proteins can be used to screen drugs or compounds that modulate the ORFX activity or expression as well as to treat disorders characterized by insufficient or excessive production of ORFX protein, for

example proliferative or differentiative disorders, or production of ORFX protein forms that have decreased or aberrant activity compared to ORFX wild type protein. In addition, the anti-ORFX antibodies of the invention can be used to detect and isolate ORFX proteins and modulate ORFX activity.

This invention further pertains to novel agents identified by the above described screening assays and uses thereof for treatments as described herein.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to ORFX proteins or have a stimulatory or inhibitory effect on, for example, ORFX expression or ORFX activity.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a ORFX protein or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc Natl Acad Sci U.S.A.* 90:6909; Erb *et al.* (1994) *Proc Natl Acad Sci U.S.A.* 91:11422; Zuckermann *et al.* (1994) *J Med Chem* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew Chem Int Ed Engl* 33:2059; Carell *et al.* (1994) *Angew Chem Int Ed Engl* 33:2061; and Gallop *et al.* (1994) *J Med Chem* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), on chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner U.S. Pat. No. 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and

Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc Natl Acad Sci U.S.A.* 87:6378-6382; Felici (1991) *J Mol Biol* 222:301-310; Ladner above.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a ORFX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the ORFX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the ORFX protein or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the test compound to preferentially bind to ORFX or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of ORFX protein, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the ORFX protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of ORFX or a biologically active portion thereof can be accomplished, for example, by determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule. As used herein, a "target molecule" is a molecule with which a ORFX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a ORFX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule

associated with the internal surface of a cell membrane or a cytoplasmic molecule. A ORFX target molecule can be a non-ORFX molecule or a ORFX protein or polypeptide of the present invention. In one embodiment, a ORFX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.*, a signal generated by binding of a compound to a membrane-bound ORFX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with ORFX.

Determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the ORFX protein to bind to or interact with a ORFX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a ORFX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a ORFX protein or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the ORFX protein or biologically active portion thereof. Binding of the test compound to the ORFX protein can be determined either directly or indirectly as described above. In one embodiment, the assay comprises contacting the ORFX protein or biologically active portion thereof with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the test compound to preferentially bind to ORFX or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting ORFX protein or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the ORFX protein or

biologically active portion thereof. Determining the ability of the test compound to modulate the activity of ORFX can be accomplished, for example, by determining the ability of the ORFX protein to bind to a ORFX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of ORFX can be accomplished by determining the ability of the ORFX protein further modulate a ORFX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting the ORFX protein or biologically active portion thereof with a known compound which binds ORFX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a ORFX protein, wherein determining the ability of the test compound to interact with a ORFX protein comprises determining the ability of the ORFX protein to preferentially bind to or modulate the activity of a ORFX target molecule.

The cell-free assays of the present invention are amenable to use of both the soluble form or the membrane-bound form of ORFX. In the case of cell-free assays comprising the membrane-bound form of ORFX, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of ORFX is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl)dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either ORFX or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to ORFX, or interaction of ORFX with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For

example, GST-ORFX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or ORFX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of ORFX binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either ORFX or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated ORFX or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with ORFX or target molecules, but which do not interfere with binding of the ORFX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or ORFX trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the ORFX or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the ORFX or target molecule.

In another embodiment, modulators of ORFX expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of ORFX mRNA or protein in the cell is determined. The level of expression of ORFX mRNA or protein in the presence of the candidate compound is compared to the level of expression of ORFX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of ORFX expression based on this comparison. For example, when expression of ORFX mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of ORFX mRNA or protein expression. Alternatively, when expression of ORFX mRNA or protein is less (statistically significantly less) in the presence of the candidate

compound than in its absence, the candidate compound is identified as an inhibitor of ORFX mRNA or protein expression. The level of ORFX mRNA or protein expression in the cells can be determined by methods described herein for detecting ORFX mRNA or protein.

In yet another aspect of the invention, the ORFX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, *e.g.*, U.S. Pat. No. 5,283,317; Zervos *et al.* (1993) Cell 72:223-232; Madura *et al.* (1993) J Biol Chem 268:12046-12054; Bartel *et al.* (1993) Biotechniques 14:920-924; Iwabuchi *et al.* (1993) Oncogene 8:1693-1696; and Brent WO94/10300), to identify other proteins that bind to or interact with ORFX ("ORFX-binding proteins" or "ORFX-bp") and modulate ORFX activity. Such ORFX-binding proteins are also likely to be involved in the propagation of signals by the ORFX proteins as, for example, upstream or downstream elements of the ORFX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for ORFX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a ORFX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with ORFX.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a

minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample.

The ORFX sequences of the present invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Pat. No. 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the ORFX sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The ORFX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:2*n*-1 (wherein $n = 1$ to 3161), as described above, can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining ORFX protein and/or nucleic acid expression as well as ORFX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant ORFX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with ORFX protein, nucleic acid expression or activity. For example, mutations in a ORFX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with ORFX protein, nucleic acid expression or activity.

Another aspect of the invention provides methods for determining ORFX protein, nucleic acid expression or ORFX activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX in clinical trials.

Use of Partial ORFX Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen

found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, that can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NOs:__ are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the ORFX sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of one or more of SEQ ID NO:2*n*-1 (where $n = 1$ to 3161), having a length of at least 20 bases, preferably at least 30 bases.

The ORFX sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or label-able probes that can be used, for example, in an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue, etc. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such ORFX probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, ORFX primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining ORFX protein and/or nucleic acid expression as well as ORFX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant ORFX expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with ORFX protein, nucleic

acid expression or activity. For example, mutations in a ORFX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with ORFX protein, nucleic acid expression or activity.

Another aspect of the invention provides methods for determining ORFX protein, nucleic acid expression or ORFX activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

Other conditions in which proliferation of cells plays a role include tumors, restenosis, psoriasis, Dupuytren's contracture, diabetic complications, Kaposi's sarcoma and rheumatoid arthritis.

An ORFX polypeptide may be used to identify an interacting polypeptide a sample or tissue. The method comprises contacting the sample or tissue with ORFX, allowing formation of a complex between the ORFX polypeptide and the interacting polypeptide, and detecting the complex, if present.

The proteins of the invention may be used to stimulate production of antibodies specifically binding the proteins. Such antibodies may be used in immunodiagnostic procedures to detect the occurrence of the protein in a sample. The proteins of the invention may be used to stimulate cell growth and cell proliferation in conditions in which such growth would be favorable. An example would be to counteract toxic side effects of chemotherapeutic agents on, for example, hematopoiesis and platelet formation, linings of the gastrointestinal tract, and hair follicles. They may also be used to stimulate new cell growth in neurological disorders including, for example, Alzheimer's disease. Alternatively, antagonistic treatments may be administered in which an antibody specifically binding the ORFX -like proteins of the invention

would abrogate the specific growth-inducing effects of the proteins. Such antibodies may be useful, for example, in the treatment of proliferative disorders including various tumors and benign hyperplasias.

Polynucleotides or oligonucleotides corresponding to any one portion of the ORFX nucleic acids of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161) may be used to detect DNA containing a corresponding ORF gene, or detect the expression of a corresponding ORFX gene, or ORFX-like gene. For example, an ORFX nucleic acid expressed in a particular cell or tissue, as noted in Table 2, can be used to identify the presence of that particular cell type.

An exemplary method for detecting the presence or absence of ORFX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting ORFX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes ORFX protein such that the presence of ORFX is detected in the biological sample. An agent for detecting ORFX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to ORFX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length ORFX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1 (wherein *n* = 1 to 3161), or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to ORFX mRNA or genomic DNA, as described above. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting ORFX protein is an antibody capable of binding to ORFX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect ORFX mRNA, protein, or genomic DNA in a biological sample *in vitro* as

well as *in vivo*. For example, *in vitro* techniques for detection of ORFX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of ORFX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of ORFX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of ORFX protein include introducing into a subject a labeled anti-ORFX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting ORFX protein, mRNA, or genomic DNA, such that the presence of ORFX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of ORFX protein, mRNA or genomic DNA in the control sample with the presence of ORFX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of ORFX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting ORFX protein or mRNA in a biological sample; means for determining the amount of ORFX in the sample; and means for comparing the amount of ORFX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect ORFX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant ORFX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with ORFX protein, nucleic acid expression or activity in, *e.g.*, proliferative or differentiative disorders such as hyperplasias, tumors, restenosis, psoriasis, Dupuytren's

contracture, diabetic complications, or rheumatoid arthritis, etc.; and glia-associated disorders such as cerebral lesions, diabetic neuropathies, cerebral edema, senile dementia, Alzheimer's disease, etc. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant ORFX expression or activity in which a test sample is obtained from a subject and ORFX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of ORFX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant ORFX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant ORFX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder, such as a proliferative disorder, differentiative disorder, glia-associated disorders, etc. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant ORFX expression or activity in which a test sample is obtained and ORFX protein or nucleic acid is detected (*e.g.*, wherein the presence of ORFX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant ORFX expression or activity.)

The methods of the invention can also be used to detect genetic lesions in a ORFX gene, thereby determining if a subject with the lesioned gene is at risk for, or suffers from, a proliferative disorder, differentiative disorder, glia-associated disorder, etc. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a ORFX-protein, or the mis-expression of the ORFX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of (1) a deletion of one or more nucleotides from a ORFX gene; (2) an addition of one or more nucleotides to a ORFX gene; (3) a substitution of one or more nucleotides of a ORFX gene, (4) a chromosomal

rearrangement of a ORFX gene; (5) an alteration in the level of a messenger RNA transcript of a ORFX gene, (6) aberrant modification of a ORFX gene, such as of the methylation pattern of the genomic DNA, (7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a ORFX gene, (8) a non-wild type level of a ORFX-protein, (9) allelic loss of a ORFX gene, and (10) inappropriate post-translational modification of a ORFX-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a ORFX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *PNAS* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the ORFX-gene (see Abravaya *et al.* (1995) *Nucl Acids Res* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a ORFX gene under conditions such that hybridization and amplification of the ORFX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli *et al.*, 1990, *Proc Natl Acad Sci USA* 87:1874-1878), transcriptional amplification system (Kwoh, *et al.*, 1989, *Proc Natl Acad Sci USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al*, 1988, *BioTechnology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a ORFX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Pat. No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in ORFX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin *et al.* (1996) *Human Mutation* 7: 244-255; Kozal *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in ORFX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin *et al.* above. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the ORFX gene and detect mutations by comparing the sequence of the sample ORFX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert (1977) *PNAS* 74:560 or Sanger (1977) *PNAS* 74:5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve *et al.*, (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publ. No. WO 94/16101; Cohen *et al.* (1996) *Adv Chromatogr* 36:127-162; and Griffin *et al.* (1993) *Appl Biochem Biotechnol* 38:147-159).

Other methods for detecting mutations in the ORFX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA

heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type ORFX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al* (1988) *Proc Natl Acad Sci USA* 85:4397; Saleeba *et al* (1992) *Methods Enzymol* 217:286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in ORFX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a ORFX sequence, *e.g.*, a wild-type ORFX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Pat. No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in ORFX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl Acad Sci USA*: 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control ORFX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting

alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA, rather than DNA, in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. See, *e.g.*, Keen *et al.* (1991) *Trends Genet* 7:5.

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). See, *e.g.*, Myers *et al* (1985) *Nature* 313:495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, *e.g.*, Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. See, *e.g.*, Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc Natl Acad. Sci USA* 86:6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. See, *e.g.*, Gasparini *et al* (1992) *Mol Cell Probes* 6:1. It is anticipated that in certain embodiments amplification may also be performed using Taq ligase

for amplification. See, *e.g.*, Barany (1991) *Proc Natl Acad Sci USA* 88:189. In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a ORFX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which ORFX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on ORFX activity (*e.g.*, ORFX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g.*, neurological, cancer-related or gestational disorders) associated with aberrant ORFX activity. In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of ORFX protein, expression of ORFX nucleic acid, or mutation content of ORFX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996, *Clin Exp Pharmacol Physiol*, 23:983-985 and Linder, 1997, *Clin Chem*, 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic

conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of ORFX protein, expression of ORFX nucleic acid, or mutation content of ORFX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a ORFX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring Clinical Efficacy

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of ORFX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied in basic drug screening and in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase ORFX gene expression, protein levels, or upregulate ORFX activity, can be monitored in clinical trials of subjects exhibiting decreased ORFX gene expression, protein levels, or downregulated ORFX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease ORFX gene expression, protein levels, or downregulate ORFX activity, can be monitored in clinical trials of subjects exhibiting increased ORFX gene expression, protein levels, or upregulated ORFX activity. In such clinical trials, the expression or activity of ORFX and, preferably, other genes that have been implicated in, for example, a proliferative or neurological disorder, can be used as a "read out" or marker of the responsiveness of a particular cell.

For example, genes, including ORFX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates ORFX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of ORFX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of ORFX or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, nucleic acid, peptidomimetic, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of a ORFX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more

post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the ORFX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the ORFX protein, mRNA, or genomic DNA in the pre-administration sample with the ORFX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of ORFX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of ORFX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant ORFX expression or activity.

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, (*i*) a ORFX polypeptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to a ORFX peptide; (*iii*) nucleic acids encoding a ORFX peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to a ORFX peptide) that are utilized to "knockout" endogenous function of a ORFX peptide by homologous recombination (see, *e.g.*, Capecchi, 1989, *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between a ORFX peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized

include, but are not limited to, a ORFX peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of a ORFX peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, etc.).

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant ORFX expression or activity, by administering to the subject an agent that modulates ORFX expression or at least one ORFX activity. Subjects at risk for a disease that is caused or contributed to by aberrant ORFX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the ORFX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of ORFX aberrancy, for example, a ORFX agonist or ORFX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

Another aspect of the invention pertains to methods of modulating ORFX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of ORFX protein activity associated with the cell. An agent that modulates ORFX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a ORFX protein, a peptide, a ORFX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more ORFX protein activity. Examples of such stimulatory agents include active ORFX protein and a nucleic acid molecule encoding ORFX that has been introduced into the cell. In another embodiment, the agent inhibits one or more ORFX protein activity. Examples of such inhibitory agents include antisense ORFX nucleic acid molecules and anti-ORFX antibodies. These modulatory methods can be performed *in vitro*

(*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a ORFX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) ORFX expression or activity. In another embodiment, the method involves administering a ORFX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant ORFX expression or activity.

Determination of the Biological Effect of a Therapeutic

In various embodiments of the present invention, suitable *in vitro* or *in vivo* assays are utilized to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Malignancies

Some ORFX polypeptides are expressed in cancerous cells (*see, e.g.*, Tables 1 and 2). Accordingly, the corresponding ORF protein is involved in the regulation of cell proliferation. Accordingly, Therapeutics of the present invention may be useful in the therapeutic or prophylactic treatment of diseases or disorders that are associated with cell hyperproliferation and/or loss of control of cell proliferation (*e.g.*, cancers, malignancies and tumors). For a review of such hyperproliferation disorders, see *e.g.*, Fishman, *et al.*, 1985. MEDICINE, 2nd ed., J.B. Lippincott Co., Philadelphia, PA.

Therapeutics of the present invention may be assayed by any method known within the art for efficacy in treating or preventing malignancies and related disorders. Such assays include,

but are not limited to, *in vitro* assays utilizing transformed cells or cells derived from the patient's tumor, as well as *in vivo* assays using animal models of cancer or malignancies. Potentially effective Therapeutics are those that, for example, inhibit the proliferation of tumor-derived or transformed cells in culture or cause a regression of tumors in animal models, in comparison to the controls.

In the practice of the present invention, once a malignancy or cancer has been shown to be amenable to treatment by modulating (*i.e.*, inhibiting, antagonizing or agonizing) activity, that cancer or malignancy may subsequently be treated or prevented by the administration of a Therapeutic that serves to modulate protein function.

Premalignant conditions

The Therapeutics of the present invention that are effective in the therapeutic or prophylactic treatment of cancer or malignancies may also be administered for the treatment of pre-malignant conditions and/or to prevent the progression of a pre-malignancy to a neoplastic or malignant state. Such prophylactic or therapeutic use is indicated in conditions known or suspected of preceding progression to neoplasia or cancer, in particular, where non-neoplastic cell growth consisting of hyperplasia, metaplasia or, most particularly, dysplasia has occurred. For a review of such abnormal cell growth see *e.g.*, Robbins & Angell, 1976. BASIC PATHOLOGY, 2nd ed., W.B. Saunders Co., Philadelphia, PA.

Hyperplasia is a form of controlled cell proliferation involving an increase in cell number in a tissue or organ, without significant alteration in its structure or function. For example, it has been demonstrated that endometrial hyperplasia often precedes endometrial cancer. Metaplasia is a form of controlled cell growth in which one type of mature or fully differentiated cell substitutes for another type of mature cell. Metaplasia may occur in epithelial or connective tissue cells. Dysplasia is generally considered a precursor of cancer, and is found mainly in the epithelia. Dysplasia is the most disorderly form of non-neoplastic cell growth, and involves a loss in individual cell uniformity and in the architectural orientation of cells. Dysplasia characteristically occurs where there exists chronic irritation or inflammation, and is often found in the cervix, respiratory passages, oral cavity, and gall bladder.

Alternatively, or in addition to the presence of abnormal cell growth characterized as hyperplasia, metaplasia, or dysplasia, the presence of one or more characteristics of a transformed or malignant phenotype displayed either *in vivo* or *in vitro* within a cell sample derived from a patient, is indicative of the desirability of prophylactic/therapeutic administration of a Therapeutic that possesses the ability to modulate activity of An aforementioned protein. Characteristics of a transformed phenotype include, but are not limited to: (*i*) morphological changes; (*ii*) looser substratum attachment; (*iii*) loss of cell-to-cell contact inhibition; (*iv*) loss of anchorage dependence; (*v*) protease release; (*vi*) increased sugar transport; (*vii*) decreased serum requirement; (*viii*) expression of fetal antigens, (*ix*) disappearance of the 250 kDal cell-surface protein, and the like. See *e.g.*, Richards, *et al.*, 1986. MOLECULAR PATHOLOGY, W.B. Saunders Co., Philadelphia, PA.

In a specific embodiment of the present invention, a patient that exhibits one or more of the following predisposing factors for malignancy is treated by administration of an effective amount of a Therapeutic: (*i*) a chromosomal translocation associated with a malignancy (*e.g.*, the Philadelphia chromosome (*bcr/abl*) for chronic myelogenous leukemia and t(14;18) for follicular lymphoma, etc.); (*ii*) familial polyposis or Gardner's syndrome (possible forerunners of colon cancer); (*iii*) monoclonal gammopathy of undetermined significance (a possible precursor of multiple myeloma) and (*iv*) a first degree kinship with persons having a cancer or pre-cancerous disease showing a Mendelian (genetic) inheritance pattern (*e.g.*, familial polyposis of the colon, Gardner's syndrome, hereditary exostosis, polyendocrine adenomatosis, Peutz-Jeghers syndrome, neurofibromatosis of Von Recklinghausen, medullary thyroid carcinoma with amyloid production and pheochromocytoma, retinoblastoma, carotid body tumor, cutaneous melanocarcinoma, intraocular melanocarcinoma, xeroderma pigmentosum, ataxia telangiectasia, Chediak-Higashi syndrome, albinism, Fanconi's aplastic anemia and Bloom's syndrome).

In another embodiment, a Therapeutic of the present invention is administered to a human patient to prevent the progression to breast, colon, lung, pancreatic, or uterine cancer, or melanoma or sarcoma.

Hyperproliferative and dysproliferative disorders

In one embodiment of the present invention, a Therapeutic is administered in the therapeutic or prophylactic treatment of hyperproliferative or benign dysproliferative disorders. The efficacy in treating or preventing hyperproliferative diseases or disorders of a Therapeutic of the present invention may be assayed by any method known within the art. Such assays include *in vitro* cell proliferation assays, *in vitro* or *in vivo* assays using animal models of hyperproliferative diseases or disorders, or the like. Potentially effective Therapeutics may, for example, promote cell proliferation in culture or cause growth or cell proliferation in animal models in comparison to controls.

Specific embodiments of the present invention are directed to the treatment or prevention of cirrhosis of the liver (a condition in which scarring has overtaken normal liver regeneration processes); treatment of keloid (hypertrophic scar) formation causing disfiguring of the skin in which the scarring process interferes with normal renewal; psoriasis (a common skin condition characterized by excessive proliferation of the skin and delay in proper cell fate determination); benign tumors; fibrocystic conditions and tissue hypertrophy (*e.g.*, benign prostatic hypertrophy).

Neurodegenerative disorders

Some ORFX proteins are found in cell types have been implicated in the deregulation of cellular maturation and apoptosis, which are both characteristic of neurodegenerative disease. Accordingly, Therapeutics of the invention, particularly but not limited to those that modulate (or supply) activity of an aforementioned protein, may be effective in treating or preventing neurodegenerative disease. Therapeutics of the present invention that modulate the activity of an aforementioned protein involved in neurodegenerative disorders can be assayed by any method known in the art for efficacy in treating or preventing such neurodegenerative diseases and disorders. Such assays include *in vitro* assays for regulated cell maturation or inhibition of apoptosis or *in vivo* assays using animal models of neurodegenerative diseases or disorders, or any of the assays described below. Potentially effective Therapeutics, for example but not by way of limitation, promote regulated cell maturation and prevent cell apoptosis in culture, or reduce neurodegeneration in animal models in comparison to controls.

Once a neurodegenerative disease or disorder has been shown to be amenable to treatment by modulation activity, that neurodegenerative disease or disorder can be treated or prevented by administration of a Therapeutic that modulates activity. Such diseases include all degenerative disorders involved with aging, especially osteoarthritis and neurodegenerative disorders.

Disorders related to organ transplantation

Some ORFX can be associated with disorders related to organ transplantation, in particular but not limited to organ rejection. Therapeutics of the invention, particularly those that modulate (or supply) activity, may be effective in treating or preventing diseases or disorders related to organ transplantation. Therapeutics of the invention (particularly Therapeutics that modulate the levels or activity of an aforementioned protein) can be assayed by any method known in the art for efficacy in treating or preventing such diseases and disorders related to organ transplantation. Such assays include *in vitro* assays for using cell culture models as described below, or *in vivo* assays using animal models of diseases and disorders related to organ transplantation, see *e.g.*, below. Potentially effective Therapeutics, for example but not by way of limitation, reduce immune rejection responses in animal models in comparison to controls.

Accordingly, once diseases and disorders related to organ transplantation are shown to be amenable to treatment by modulation of activity, such diseases or disorders can be treated or prevented by administration of a Therapeutic that modulates activity.

Cardiovascular Disease

GENX has been implicated in cardiovascular disorders, including in atherosclerotic plaque formation. Diseases such as cardiovascular disease, including cerebral thrombosis or hemorrhage, ischemic heart or renal disease, peripheral vascular disease, or thrombosis of other major vessel, and other diseases, including diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage disease, systemic lupus erythematosus, homocysteinemia, and familial protein or lipid processing diseases, and the like, are either directly or indirectly associated with atherosclerosis. Accordingly, Therapeutics of the invention, particularly those that modulate (or supply) activity or formation may be effective in treating or preventing

atherosclerosis-associated diseases or disorders. Therapeutics of the invention (particularly Therapeutics that modulate the levels or activity) can be assayed by any method known in the art, including those described below, for efficacy in treating or preventing such diseases and disorders.

A vast array of animal and cell culture models exist for processes involved in atherosclerosis. A limited and non-exclusive list of animal models includes knockout mice for premature atherosclerosis (Kurabayashi and Yazaki, 1996, Int. Angiol. 15: 187-194), transgenic mouse models of atherosclerosis (Kappel *et al.*, 1994, FASEB J. 8: 583-592), antisense oligonucleotide treatment of animal models (Callow, 1995, Curr. Opin. Cardiol. 10: 569-576), transgenic rabbit models for atherosclerosis (Taylor, 1997, Ann. N.Y. Acad. Sci 811: 146-152), hypercholesterolemic animal models (Rosenfeld, 1996, Diabetes Res. Clin. Pract. 30 Suppl.: 1-11), hyperlipidemic mice (Paigen *et al.*, 1994, Curr. Opin. Lipidol. 5: 258-264), and inhibition of lipoxygenase in animals (Sigal *et al.*, 1994, Ann. N.Y. Acad. Sci. 714: 211-224). In addition, *in vitro* cell models include but are not limited to monocytes exposed to low density lipoprotein (Frostegard *et al.*, 1996, Atherosclerosis 121: 93-103), cloned vascular smooth muscle cells (Suttles *et al.*, 1995, Exp. Cell Res. 218: 331-338), endothelial cell-derived chemoattractant exposed T cells (Katz *et al.*, 1994, J. Leukoc. Biol. 55: 567-573), cultured human aortic endothelial cells (Farber *et al.*, 1992, Am. J. Physiol. 262: H1088-1085), and foam cell cultures (Libby *et al.*, 1996, Curr Opin Lipidol 7: 330-335). Potentially effective Therapeutics, for example but not by way of limitation, reduce foam cell formation in cell culture models, or reduce atherosclerotic plaque formation in hypercholesterolemic mouse models of atherosclerosis in comparison to controls.

Accordingly, once an atherosclerosis-associated disease or disorder has been shown to be amenable to treatment by modulation of activity or formation, that disease or disorder can be treated or prevented by administration of a Therapeutic that modulates activity.

Cytokine and Cell Proliferation/Differentiation Activity

A GENX protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered

to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods: Assays for T-cell or thymocyte proliferation include without limitation those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Ed by Coligan *et al.*, Greene Publishing Associates and Wiley-Interscience (Chapter 3 and Chapter 7); Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Bertagnoili *et al.*, *J Immunol* 145:1706-1712, 1990; Bertagnolli *et al.*, *Cell Immunol* 133:327-341, 1991; Bertagnolli, *et al.*, *J Immunol* 149:3778-3783, 1992; Bowman *et al.*, *J Immunol* 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described by Kruisbeek and Shevach, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1, pp. 3.12.1-14, John Wiley and Sons, Toronto 1994; and by Schreiber, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan eds. Vol 1 pp. 6.8.1-8, John Wiley and Sons, Toronto 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described by Bottomly *et al.*, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto 1991; deVries *et al.*, *J Exp Med* 173:1205-1211, 1991; Moreau *et al.*, *Nature* 336:690-692, 1988; Greenberger *et al.*, *Proc Natl Acad Sci U.S.A.* 80:2931-2938, 1983; Nordan, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.6.1-5, John Wiley and Sons, Toronto 1991; Smith *et al.*, *Proc Natl Acad Sci U.S.A.* 83:1857-1861, 1986; Measurement of human Interleukin 11-Bennett, *et* al. In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto 1991; Ciarletta, *et al.*, In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and

cytokine production) include, without limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds., Greene Publishing Associates and Wiley-Interscience (Chapter 3Chapter 6, Chapter 7); Weinberger *et al.*, *Proc Natl Acad Sci USA* 77:6091-6095, 1980; Weinberger *et al.*, *Eur J Immun* 11:405-411, 1981; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988.

Immune Stimulating or Suppressing Activity

A GENX protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), *e.g.*, in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by vital (*e.g.*, HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by vital, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, Leishmania species., malaria species. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, *i.e.*, in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be possible to immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response.

The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or energy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon re-exposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7), *e.g.*, preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (*e.g.*, B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to energize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the

immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, Science 257:789-792 (1992) and Turka *et al.*, Proc Natl Acad Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., FUNDAMENTAL IMMUNOLOGY, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and auto-antibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of auto-antibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., FUNDAMENTAL IMMUNOLOGY, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic vital diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the *in vitro* activated T cells into the

patient. Another method of enhancing anti-vital immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells *in vivo*.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (*e.g.*, sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (*e.g.*, a cytoplasmic-domain truncated portion) of an MHC class I α chain protein and β_2 microglobulin protein or an MHC class II a chain protein and an MHC class II β chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (*e.g.*, B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor

specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods: Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 7); Herrmann *et al.*, *Proc Natl Acad Sci USA* 78:2488-2492, 1981; Herrmann *et al.*, *J Immunol* 128:1968-1974, 1982; Handa *et al.*, *J Immunol* 135:1564-1572, 1985; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988; Herrmann *et al.*, *Proc Natl Acad Sci USA* 78:2488-2492, 1981; Herrmann *et al.*, *J Immunol* 128:1968-1974, 1982; Handa *et al.*, *J Immunol* 135:1564-1572, 1985; Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Bowman *et al.*, *J Virology* 61:1992-1998; Takai *et al.*, *J Immunol* 140:508-512, 1988; Bertagnolli *et al.*, *Cell Immunol* 133:327-341, 1991; Brown *et al.*, *J Immunol* 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, *J Immunol* 144:3028-3033, 1990; and Mond and Brunswick In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, (eds.) Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described In: CURRENT PROTOCOLS IN IMMUNOLOGY. Coligan *et al.*, eds. Greene Publishing Associates and Wiley-Interscience (Chapter 3, Chapter 7); Takai *et al.*, *J Immunol* 137:3494-3500, 1986; Takai *et al.*, *J Immunol* 140:508-512, 1988; Bertagnolli *et al.*, *J Immunol* 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery *et al.*, *J Immunol* 134:536-544, 1995; Inaba *et al.*, *J Exp Med* 173:549-559, 1991; Macatonia *et al.*, *J Immunol* 154:5071-5079, 1995; Porgador *et al.*, *J Exp Med* 182:255-260, 1995; Nair *et al.*, *J Virol* 67:4062-4069, 1993; Huang *et al.*, *Science* 264:961-965, 1994; Macatonia *et al.*, *J Exp Med* 169:1255-1264, 1989; Bhardwaj *et al.*, *J Clin Investig* 94:797-807, 1994; and Inaba *et al.*, *J Exp Med* 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz *et al.*, *Cytometry* 13:795-808, 1992; Gorczyca *et al.*, *Leukemia* 7:659-670, 1993; Gorczyca *et al.*, *Cancer Res* 53:1945-1951, 1993; Itoh *et al.*, *Cell* 66:233-243, 1991; Zacharchuk, *J Immunol* 145:4037-4045, 1990; Zamai *et al.*, *Cytometry* 14:891-897, 1993; Gorczyca *et al.*, *Internat J Oncol* 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica *et al.*, *Blood* 84:111-117, 1994; Fine *et al.*, *Cell Immunol* 155: 111-122, 1994; Galy *et al.*, *Blood* 85:2770-2778, 1995; Toki *et al.*, *Proc Nat Acad Sci USA* 88:7548-7551, 1991.

Hematopoiesis Regulating Activity

A GENX protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, *e.g.* in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (*i.e.*, traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (*i.e.*, in conjunction with bone marrow

transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson *et al. Cellular Biology* 15:141-151, 1995; Keller *et al.*, *Mol. Cell. Biol.* 13:473-486, 1993; McClanahan *et al.*, *Blood* 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* (eds.) Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y 1994; Hirayama *et al.*, *Proc Natl Acad Sci USA* 89:5907-5911, 1992; McNiece and Briddeli, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* (eds.) Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben *et al.*, *Exp Hematol* 22:353-359, 1994; Ploemacher, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.* eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Spoonceret *al.*, In: CULTURE OF HEMATOPOIETIC CELLS. Freshhey, *et al.*, (eds.) Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Sutherland, In: CULTURE OF HEMATOPOIETIC CELLS. Freshney, *et al.*, (eds.) Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

Tissue Growth Activity

A GENX protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation

induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendonitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a career as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, *i.e.* for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein

may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, EPIDERMAL WOUND HEALING, pp. 71-112 (Maibach and Rovee, eds.), Year Book Medical

Publishers, Inc., Chicago, as modified by Eaglstein and Menz, *J. Invest. Dermatol* 71:382-84 (1978).

Activin/Inhibin Activity

A GENX protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin a family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin-b group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale *et al.*, *Endocrinology* 91:562-572, 1972; Ling *et al.*, *Nature* 321:779-782, 1986; Vale *et al.*, *Nature* 321:776-779, 1986; Mason *et al.*, *Nature* 318:659-663, 1985; Forage *et al.*, *Proc Natl Acad Sci USA* 83:3091-3095, 1986.

Chemotactic/Chemokinetic Activity

A protein of the present invention may have chemotactic or chemokinetic activity (*e.g.*, act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example,

attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Coligan *et al.*, eds. (Chapter 6.12, MEASUREMENT OF ALPHA AND BETA CHEMOKINES 6.12.1-6.12.28); Taub *et al. J Clin Invest* 95:1370-1376, 1995; Lind *et al. APMIS* 103:140-146, 1995; Muller *et al., Eur J Immunol* 25: 1744-1748; Gruber*et al. J Immunol* 152:5860-5867, 1994; Johnston *et al., J Immunol* 153: 1762-1768, 1994.

Hemostatic and Thrombolytic Activity

A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (*e.g.*, stroke).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet *et al.*, *J. Clin. Pharmacol.* 26:131-140, 1986; Burdick *et al.*, *Thrombosis Res.*

45:413-419, 1987; Humphrey *et al.*, *Fibrinolysis* 5:71-79 (1991); Schaub, *Prostaglandins* 35:467-474, 1988.

Receptor/Ligand Activity

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell—cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: CURRENT PROTOCOLS IN IMMUNOLOGY, Ed by Coligan, *et al.*, Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai *et al.*, *Proc Natl Acad Sci USA* 84:6864-6868, 1987; Bierer *et al.*, *J. Exp. Med.* 168:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169:149-160 1989; Stoltenborg *et al.*, *J Immunol Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670, 1995.

Anti-Inflammatory Activity

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell—cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute

conditions), including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

Tumor Inhibition Activity

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth.

Other Activities

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting

deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

Neural disorders in general include Parkinson's disease, Alzheimer's disease, Huntington's disease, multiple sclerosis, amyotrophic lateral sclerosis (ALS), peripheral neuropathy, tumors of the nervous system, exposure to neurotoxins, acute brain injury, peripheral nerve trauma or injury, and other neuropathies, epilepsy, and/or tremors.

EQUIVALENTS

From the foregoing detailed description of the specific embodiments of the invention, it should be apparent that particular novel compositions and methods involving nucleic acids, polypeptides, antibodies, detection and treatment have been described. Although these particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims that follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made as a matter of routine for a person of ordinary skill in the art to the invention without departing from the spirit and scope of the invention as defined by the claims. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

THIS PAGE INTENTIONALLY LEFT BLANK

Table 1

ORF#	Internal Identification Number	Protein similarity	Protein domain	Protein Classification	Cells or Tissues in which Gene is Expressed
1	13076366 (1, 2)	Novel Protein sim. GBank gi|4691395|emb|CAB41562.1| - (AL049727) putative large secreted protein [Streptomyces coelicolor]		UNCLASSIFIED	264636
2	80248091 (3, 4)	Novel Protein sim. GBank gi|2829506|sp|P71559|SUCC_MYCTU - SUCCINYL-COA SYNTHETASE BETA CHAIN (SCS-BETA)	Contains protein domain (PF00549) - CoA-ligases	UNCLASSIFIED	264907, 264600, 264602, 264762, 264769, 264689, 264638, 264567
3	80415924 (5, 6)			UNCLASSIFIED	264910, 264604, 264634, 264905, 264636, 264691, 264907, 264692, 264629
4	82018837 (7, 8)			UNCLASSIFIED	264908, 264909, 264760, 264628, 264635
5	79970035 (9, 10)			UNCLASSIFIED	22279002, 264563
6	79842462 (11, 12)		Contains protein domain (PF00127) - Copper binding proteins, plastocyanin/azurin family	UNCLASSIFIED	264908
7	85515576 (13, 14)	Novel Protein sim. GBank gi|4415926|gb|AAD20157| - (AC006282) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	20281099, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264511, 265006, 264512, 265009, 264910, 264595, 264596, 264758, 264603, 264604, 264760, 264762, 264683, 264766, 264767, 264689, 35695917, 264690, 264692, 264693, 33657109, 264628, 264629, 35696423, 55811576, 35695855, 264630, 264631, 264632, 264634, 264636, 264637, 264638, 264639, 18108385, 264563, 264564, 264566, 264486
8	56924278 (15, 16)	Novel Protein sim. GBank gi|585562|sp|Q06458|NIRB_KLEPN - NITRITE REDUCTASE (NAD(P)H) LARGE SUBUNIT		reductase	264907
9	79394457 (17, 18)			UNCLASSIFIED	265007, 265019, 263972
10	79556459 (19, 20)			UNCLASSIFIED	264906
11	20414027 (21, 22)				264605
12	94141210 (23, 24)	Novel Protein sim. GBank gi|3878145|emb|CAA99871| - (Z75543) similar to potassium channel protein [Caenorhabditis elegans]		misc_channel	264259, 265007, 83373044
13	20750551 (25, 26)			UNCLASSIFIED	264556, 264557, 264564
14	95105114 (27, 28)	Novel Protein sim. GBank gi|2832781|emb|CAA12645| - (AJ225805) inward potassium channel alpha subunit [Egeria densa]	Contains protein domain (PF00023) - Ank repeat	potassium_channel	35696286, 35696052, 264510, 35695917, 264691, 264628, 35696423, 264555, 264558, 264559, 83373044
15	20458307 (29, 30)	Novel Protein sim. GBank gi|1710791|sp|Q10234|RT05_SCHPO - PROBABLE MITOCHONDRIAL 40S RIBOSOMAL PROTEIN S5	Contains protein domain (PF00333) - Ribosomal protein S5	ribosomalprot	264604
16	20760356 (31, 32)				264555

17	20292744 (33, 34)	Novel Protein sim. GBank gi|1174884|sp|P44391|URE1_HAEIN - UREASE ALPHA SUBUNIT (UREA AMIDOHYDROLASE)	Contains protein domain (PF00449) - Urease		264600
18	80246804 (35, 36)	Novel Protein sim. GBank gi|2281102 (AC002333) - SF16 isolog [Arabidopsis thaliana]			29331827, 264555, 264557, 264638, 264558
19	80076624 (37, 38)			UNCLASSIFIED	22278996, 264907, 264910, 264600, 264693
20	20724558 (39, 40)	Novel Protein sim. GBank gi|2506112|sp|P43672|UUP_ECOLI - ABC TRANSPORTER ATP-BINDING PROTEIN UUP		transport	264602
21	80417554 (41, 42)	Novel Protein sim. GBank gi|1730203|sp|P50442|GATM_RAT - GLYCINE AMIDINOTRANSFERASE PRECURSOR (L-ARGININE:GLYCINE AMIDINOTRANSFERASE) (TRANSAMIDINASE) (AT)		UNCLASSIFIED	22278995, 264906, 265008, 265010, 265011, 264602, 264605, 264766, 264688, 21906764, 264691, 18108376, 264636, 18108387, 264486
22	11705858 (43, 44)				264685
23	80419176 (45, 46)	Novel Protein sim. GBank gi|1877329|emb|CAB07077| - (Z92771) fadE25 [Mycobacterium tuberculosis]	Contains protein domain (PF00441) - Acyl-CoA dehydrogenase	dehydrogenase	264488, 264907, 264909, 264600, 264602, 264603, 264605, 264682, 264766, 32833986, 264636, 264486
24	20291697 (47, 48)				264600
25	80253774 (49, 50)				264593
26	80255394 (51, 52)			UNCLASSIFIED	22278996, 56182435, 265018, 264566
27	80235795 (53, 54)	Novel Protein sim. GBank gi|4808369|emb|CAB42783.1| - (AL049841) putative 30S ribosomal protein S14 [Streptomyces coelicolor]	Contains protein domain (PF00253) - Ribosomal protein S14p/S29e	ribosomalprot	18108370, 35696423, 264635, 264555
28	79483561 (55, 56)			UNCLASSIFIED	264638
29	82448765 (57, 58)	Novel Protein sim. GBank gi|3122290|sp|O08333|K6PF_STRCO - 6-PHOSPHOFRUCTOKINASE (PHOSPHOFRUCTOKINASE) (PHOSPHOHEXOKINASE) (ATP-PFK)	Contains protein domain (PF00365) - Phosphofructokinase	kinase	264601, 264762, 264766, 264769, 264636
30	79199333 (59, 60)			UNCLASSIFIED	264908, 265019, 264687, 21906764, 21906766
31	19848158 (61, 62)			UNCLASSIFIED	264534
32	82449495 (63, 64)	Novel Protein sim. GBank gi|3560504 (AF027770) - unknown [Mycobacterium smegmatis]		UNCLASSIFIED	264905, 264605, 264762, 264766, 264687, 264689
33	79582628 (65, 66)	Novel Protein sim. GBank gi|2129003|pir||G64507 - hypothetical protein MJ1665 - Methanococcus jannaschii		UNCLASSIFIED	264687
34	87467657 (67, 68)			UNCLASSIFIED	60432289, 264600, 264602, 264760, 18108357, 264769, 265020, 264691
35	95005170 (69, 70)	Novel Protein sim. GBank gi|5420387|emb|CAB46679.1| - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	264600, 264687, 264558, 264639
36	19642042 (71, 72)	Novel Protein sim. GBank gi|3287739|sp|P73538|BIOB_SYNY3 - BIOTIN SYNTHASE (BIOTIN SYNTHETASE)		synthase	264566
37	20369215 (73, 74)	Novel Protein sim. GBank gi|2313134|gb|AAD07126.1| - (AE000527) delta-1-pyrroline-5-carboxylate dehydrogenase [Helicobacter pylori 26695]		dehydrogenase	264603

38	20466334 (75, 76)	Novel Protein sim. GBank gi|3805970|emb|CAA06231| - (AJ004933) periplasmic nitrate reductase, large subunit [Rhodopseudomonas sp.]		reductase	264605
39	94300715 (77, 78)	Novel Protein sim. GBank gi|1929449 (L63543) - endodermin [Xenopus laevis]	Contains protein domain (PF00207) - Alpha-2-macroglobulin family	complement	264905, 264906, 264907, 66712502, 264908, 264909, 264511, 265009, 264910, 55812038, 264758, 265011, 264762, 264682, 264763, 264764, 264766, 265022, 264693, 264628, 264631, 264634, 264635, 264555, 264638, 18108381, 264558, 18108385, 264482
40	20635625 (79, 80)			UNCLASSIFIED	264592
41	80023287 (81, 82)	Novel Protein sim. GBank gi|854065|emb|CAA58337| - (X83413) U88 [Human herpesvirus 6]			264591, 35695917
42	20724566 (83, 84)			UNCLASSIFIED	264602
43	20467069 (85, 86)	Novel Protein sim. GBank gi|3820584 (AF086791) - carbamoylphosphate synthetase large subunit [Zymomonas mobilis]		synthase	264605
44	13085297 (87, 88)	Novel Protein sim. GBank gi|2494764|sp|Q50729|GUAA_MYCTU - GMP SYNTHASE (GLUTAMINE-HYDROLYZING) (GLUTAMINE AMIDOTRANSFERASE) (GMP SYNTHETASE)	Contains protein domain (PF00958) - GMP synthase C terminal domain	synthase	264769, 264636
45	39384711 (89, 90)	Novel Protein sim. GBank gi|1881738 (U89688) - myosin-I binding protein Acan125 [Acanthamoeba castellanii]		UNCLASSIFIED	264769, 264510, 264508
46	95003398 (91, 92)			ngf	264566
47	11698624 (93, 94)			UNCLASSIFIED	264689
48	79407218 (95, 96)				18108385, 264635, 264828
49	21659844 (97, 98)			UNCLASSIFIED	264603
50	80503996 (99, 100)				264508, 264603, 264769, 264689, 264636, 264558, 264486
51	80255569 (101, 102)	Novel Protein sim. GBank gi|3411177 (AF076240) - MocC [Rhizobium leguminosarum bv. viciae]		UNCLASSIFIED	264593, 18108387
52	79208528 (103, 104)	Novel Protein sim. GBank gi|3914992|sp|Q26264|SM41_HEMPU - 41 KD SPICULE MATRIX PROTEIN PRECURSOR (HSM41) (HPSMC)		struct	264634
53	36996970 (105, 106)	Novel Protein sim. GBank gi|3980411 (AC004561) - putative proline-rich protein [Arabidopsis thaliana]		UNCLASSIFIED	264762
54	79570897 (107, 108)			UNCLASSIFIED	264630, 264909, 264766
55	80202703 (109, 110)	Novel Protein sim. GBank gi|1633572 (U52064) - Herpesvirus saimiri ORF73 homolog [Kaposi's sarcoma-associated herpes-like virus]			29331824, 264102, 265018, 18108376
56	8758408 (111, 112)	Novel Protein sim. GBank gi|4321580|gb|AAD15785| - (AF050114) alginate lyase [Pseudomonas sp. W7]			264604
57	11223386 (113, 114)		Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264557

58	91227506 (115, 116)	Novel Protein sim. GBank gi|5816074|gb|AAD45616.1|AF06194 - (AF061943) protate-derived STE20-like kinase PSK [Homo sapiens]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	56182575, 264259, 60432049, 35696052, 66712502, 264909, 265008, 265010, 265011, 264681, 29148784, 35695917, 60170615, 264691, 264692, 264693, 18108374, 35696423, 56182323, 60432113
59	80077371 (117, 118)	Novel Protein sim. GBank gi|1172920|sp|P45830|RFE_MYCLE - PUTATIVE UNDECAPRENYL-PHOSPHATE ALPHA-N-ACETYLGLUCOSAMINYLTRANSFERASE	Contains protein domain (PF00953) - Glycosyl transferase	transferase	264600, 264689, 264638
60	12958341 (119, 120)				264689
61	80426806 (121, 122)	Novel Protein sim. GBank gi|1710216 (U79260) - unknown [Homo sapiens]		glycoprotein	264766
62	13504966 (123, 124)				264630
63	16474553 (125, 126)			UNCLASSIFIED	265019
64	20724578 (127, 128)	Novel Protein sim. GBank gi|420945|pir||A47041 - transposase homolog (insertion element ISAE1) - Alcaligenes eutrophus		UNCLASSIFIED	264602
65	79326308 (129, 130)	Novel Protein sim. GBank gi|3122312|sp|O06134|KPYK_MYCTU - PYRUVATE KINASE (PK)	Contains protein domain (PF00224) - Pyruvate kinase	kinase	264563
66	46854384 (131, 132)	Novel Protein sim. GBank gi|3928723|emb|CAA22219| - (AL034355) putative ABC transporter [Streptomyces coelicolor]		transport	22278996, 264558
67	78952543 (133, 134)	Novel Protein sim. GBank gi|231985|sp|P30234|DHA_MYCTU - ALANINE DEHYDROGENASE (40 KD ANTIGEN)		dehydrogenase	265021
68	79817382 (135, 136)				264909
69	79841764 (137, 138)			UNCLASSIFIED	264908
70	79871329 (139, 140)				264906, 264908
71	65897456 (141, 142)			UNCLASSIFIED	264602, 265021
72	87734977 (143, 144)	Novel Protein sim. GBank gi|4415926|gb|AAD20157| - (AC006282) unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	264488, 264905, 264906, 264907, 264908, 264511, 265008, 264910, 264758, 87168474, 264682, 264766, 264686, 264689, 35695917, 265021, 60170615, 264691, 33657023, 264692, 264693, 264629, 264631, 264639, 22279000
73	80025241 (145, 146)			UNCLASSIFIED	60424179, 264508, 264908, 265007, 264603, 264687, 264689, 264692, 18108387
74	20377410 (147, 148)			UNCLASSIFIED	264605
75	11819032 (149, 150)	Novel Protein sim. GBank gi|2853098|emb|CAA16914| - (AL021767) vacuolar protein sorting [Schizosaccharomyces pombe]		UNCLASSIFIED	264689
76	95105303 (151, 152)	Novel Protein sim. GBank gi|4468811|emb|CAB38212| - (AL035601) putative protein [Arabidopsis thaliana]		UNCLASSIFIED	83373044, 264906, 264557
77	10144718 (153, 154)	Novel Protein sim. GBank gi|854085|emb|CAA58337| - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264563
78	8758258 (155, 156)			UNCLASSIFIED	264604

79	94140190 (157, 158)	Novel Protein sim. GBank gi|5689453|dbj|BAA83010.1| - (AB028981) KIAA1058 protein [Homo sapiens]	Contains protein domain (PF00169) - PH domain		35696286, 22278998, 29331822, 29331824, 29331825, 29331827, 264905, 264906, 264907, 66712502, 264908, 264909, 265008, 265009, 264910, 60170831, 55812038, 33109954, 265017, 265018, 264288, 264768, 56181562, 21906765, 21906769, 29148784, 265020, 264690, 264691, 264692, 264693, 60431528, 35696423, 264631, 264632, 264634, 264636, 264639, 83373044, 264564, 264566, 264567
80	82314840 (159, 160)			UNCLASSIFIED	264769, 264601, 265006, 264910, 264604, 264605, 264634, 264635, 264905, 264762, 264637, 264592, 264628, 264907, 264691, 264908, 264567, 264909, 264766
81	20467247 (161, 162)	Novel Protein sim. GBank gi|1723442|sp|Q10258|YD2A_SCHPO - HYPOTHETICAL 69.0 KD PROTEIN C56F8.10 IN CHROMOSOME I		reductase	264605
82	16331388 (163, 164)	Novel Protein sim. GBank gi|2895866 (AF045770) - methylmalonate semi-aldehyde dehydrogenase [Oryza sativa]		dehydrogenase	264567
83	94741180 (165, 166)	Novel Protein sim. GBank gi|3402673 (AC004697) - unknown protein [Arabidopsis thaliana]		UNCLASSIFIED	264488, 264508, 264509, 264905, 264908, 264909, 264511, 264591, 264593, 264594, 264595, 264596, 264758, 264603, 264760, 264681, 18108351, 264762, 264682, 264764, 264684, 264766, 264686, 264632, 264637, 264557, 264638, 264639, 18108385, 264566
84	80355375 (167, 168)	Novel Protein sim. GBank gi|1173364|sp|P45380|SAT1_RAT - SULFATE ANION TRANSPORTER 1 (CANALICULAR SULFATE TRANSPORTER) (SULFATE/CARBONATE ANTIPORTER)		transport	264508, 264906, 264907, 264908, 264909, 264910, 264760, 264763, 264764, 264766, 264768, 264769, 35695855, 264636, 264637
85	80499600 (169, 170)	Novel Protein sim. GBank gi|2120998|pir||S70682 - glycosyltransferase homolog - Bordetella pertussis		transferase	264605, 264762, 264687, 264769, 18108374, 264636, 264486
86	39559043 (171, 172)	Novel Protein sim. GBank gi|3256023|emb|CAA17228.1| - (AL021897) hypothetical protein Rv1112 [Mycobacterium tuberculosis]			264910
87	13856808 (173, 174)			UNCLASSIFIED	264093

88	95344718 (175, 176)	Novel Protein sim. GBank gi|559703|dbj|BAA07552| - (D38549) ha1025 is new [Homo sapiens]			52644507, 52646365, 18108398, 65274572, 56182575, 56994075, 35696286, 22278997, 22278998, 22278999, 264092, 264093, 264094, 264095, 264259, 29331822, 29331824, 56182181, 66714117, 29331825, 29331826, 60432289, 29331827, 29331828, 35696052, 33656970, 264105, 264508, 264905, 264906, 264907, 264908, 29331830, 66712502, 52644045, 56182435, 265007, 265009, 264910, 60170831, 264592, 60431735, 60433356, 33657402, 264757, 60433438, 55812038, 264758, 21906754, 52646317, 33109954, 52644296, 87168474, 265011, 87168559, 264601, 265017, 265018, 264604, 265019, 264448, 264369, 264288, 264766, 52644229, 21906766, 21906767, 21906768, 21906769, 55811957, 35695917, 265020, 265021, 265022, 60170615, 52644150, 33657023, 65274620, 33657109, 27486261, 27486264, 33657349, 35695763, 264628, 263972, 18108374, 55810764, 35696423, 55811576, 65274791, 35695855, 60431850, 264636, 52644332, 56182323, 60170394, 63373044, 18108385, 18108387, 18108388, 56526486, 87168518, 60432113, 22279000, 22279002, 264482, 264564; 264486
89	80077389 (177, 178)	Novel Protein sim. GBank gi|1710383|sp|P46352|RIPX_BACSU - PROBABLE INTEGRASE/RECOMBINASE RIPX			264600
90	82115999 (179, 180)			UNCLASSIFIED	264760
91	78906950 (181, 182)	Novel Protein sim. GBank gi|2499891|sp|P76403|YEGQ_ECOLI - PUTATIVE PROTEASE IN BAER-OGRK INTERGENIC REGION		protease	265006
92	79554871 (183, 184)	Novel Protein sim. GBank gi|3367754|emb|CAA20079| - (AL031155) hypothetical protein SC3A7.16c [Streptomyces coelicolor]		UNCLASSIFIED	264691
93	80496778 (185, 186)	Novel Protein sim. GBank gi|2895095 (AF011337) - putative E1-E2 ATPase [Mus musculus]		ATPase_associated	264907, 264908, 264910, 265009, 264605, 264769
94	79646649 (187, 188)	Novel Protein sim. GBank gi|1171919|sp|P46920|OPUA_BACSU - GLYCINE BETAINE TRANSPORT ATP-BINDING PROTEIN OPUAA	Contains protein domain (PF00571) - CBS domain	transport	264906
95	11090238 (189, 190)				264594

96	94322125 (191, 192)	Novel Protein sim. GBank gi|4589560|dbj|BAA76802.1| - (AB023175) KIAA0958 protein [Homo sapiens]		UNCLASSIFIED	22278995, 22278999, 264259, 29331822, 29331826, 35696052, 29146499, 264509, 264906, 264907, 264908, 264909, 265007, 265008, 264910, 265009, 264593, 265010, 265017, 264604, 265019, 18108351, 264288, 264766, 264768, 264769, 21906765, 21906767, 21906769, 265020, 264692, 33657182, 35695763, 264628, 264629, 18108379, 264631, 264636, 18108381, 264559, 18108382, 83373044, 22279002
97	79605200 (193, 194)	Novel Protein sim. GBank gi|4583559|emb|CAB40388.1| - (AJ005255) OxyR [Erwinia chrysanthemi]		UNCLASSIFIED	264508
98	79427000 (195, 196)	Novel Protein sim. GBank gi|1001693|dbj|BAA10430| - (D64002) hypothetical protein [Synechocystis sp.]		UNCLASSIFIED	264909
99	20466524 (197, 198)	Novel Protein sim. GBank gi|1169479|sp|P43925|EFG_HAEIN - ELONGATION FACTOR G (EF-G)		UNCLASSIFIED	264605
100	79640113 (199, 200)			UNCLASSIFIED	264693
101	80203298 (201, 202)	Novel Protein sim. GBank gi|480897|pir||S37485 - gene msg1 protein - mouse		UNCLASSIFIED	265020, 264102, 263972
102	20467259 (203, 204)	Novel Protein sim. GBank gi|2894166|emb|CAA11773.1| - (AJ223998) PCZA361.18 [Amycolatopsis orientalis]		synthase	264605
103	20466368 (205, 206)	Novel Protein sim. GBank gi|1731040|sp|P54509|YQHH_BACSU - HYPOTHETICAL HELICASE IN SINI-GCVT INTERGENIC REGION	Contains protein domain (PF00271) - Helicases conserved C-terminal domain	helicase	264605
104	80247572 (207, 208)	Novel Protein sim. GBank gi|854065|emb|CAA58337| - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264591, 264595, 264602
105	79605206 (209, 210)	Novel Protein sim. GBank gi|1685117 (U70770) - furrowed [Drosophila melanogaster]	Contains protein domain (PF00084) - Sushi domain (SCR repeat)	complement	264508
106	28382058 (211, 212)	Novel Protein sim. GBank gi|1705505|sp|P54729|BS4_MOUSE - BS4 PROTEIN	Contains protein domain (PF00627) - UBA domain	UNCLASSIFIED	264511, 265009
107	80057791 (213, 214)	Novel Protein sim. GBank gi|4887229|gb|AAD32244.1|AF15075 - (AF150755) microtubule-actin crosslinking factor [Mus musculus]		ATPase_associated	29331824, 264591, 21906754, 265019
108	80237936 (215, 216)	Novel Protein sim. GBank gi|2635771|emb|CAB15264| - (Z99120) similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]	Contains protein domain (PF00005) - ABC transporter	transport	18108374, 35695917, 22278996, 264113, 264600, 264602, 264603, 265017, 264910, 264906, 264636, 264766
109	95194148 (217, 218)	Novel Protein sim. GBank gi|2330791|emb|CAB11265| - (Z98601) carboxypeptidase s precursor [Schizosaccharomyces pombe]		UNCLASSIFIED	264758, 264603, 264630, 264636, 264637
110	79582823 (219, 220)				264687
111	39565458 (221, 222)				264564
112	79856038 (223, 224)			UNCLASSIFIED	264908
113	17959439 (225, 226)			UNCLASSIFIED	265007
114	80502101 (227, 228)			UNCLASSIFIED	264769

115	80251003 (229, 230)	Novel Protein sim. GBank gi|2246532 (U93872) - ORF 73, contains large complex repeat CR 73 [Kaposi's sarcoma-associated herpesvirus]		UNCLASSIFIED	52645156, 52645080, 33656970, 264592, 21908754, 27486264, 18108379, 35696423, 264635, 52644332, 18108382
116	81298689 (231, 232)				264905, 264906, 264907, 264908, 264909, 264910, 264758, 265010, 264763, 264682, 264764, 264766, 264685, 264686, 264768, 264769, 33657023, 264693, 33657109, 264628, 18108374, 264631, 264632, 264634, 264636, 264637, 264638, 264639, 56526486, 264565, 264566
117	79636695 (233, 234)				264639, 264693
118	80222170 (235, 236)		Contains protein domain (PF00170) - bZIP transcription factor	UNCLASSIFIED	263974
119	91013071 (237, 238)	Novel Protein sim. GBank gi|732526 (U22327) - alpha2(IV) collagen [Caenorhabditis elegans]		UNCLASSIFIED	22278996, 29331824, 60432289, 265007, 60433438, 264603, 264605, 18108351, 264769, 264689, 265020, 264534, 27486261, 264558, 83373044, 18108385, 264564
120	8756491 (239, 240)	Novel Protein sim. GBank gi|2131219|pir||S50157 - cyclin-dependent kinase chain SRB10 - yeast (Saccharomyces cerevisiae)		kinase	264603
121	80026153 (241, 242)				264595
122	20457620 (243, 244)	Novel Protein sim. GBank gi|2052147|emb|CAB08137| - (Z94752) ksgA [Mycobacterium tuberculosis]	Contains protein domain (PF00398) - Ribosomal RNA adenine dimethylases	transferase	264605
123	8758278 (245, 246)				264604
124	79104017 (247, 248)	Novel Protein sim. GBank gi|2833385|sp|Q43134|UGST_SORBI - GRANULE-BOUND GLYCOGEN (STARCH) SYNTHASE PRECURSOR		synthase	18108394, 18108397, 265006, 265007, 265008, 265010, 265011, 18108355, 18108379, 18108380, 18108384
125	87797986 (249, 250)	Novel Protein sim. GBank gi|475542 (U08255) - glutamate receptor delta-1 subunit [Rattus norvegicus]	Contains protein domain (PF00060) - Ligand-gated ion channel	misc_channel	264508, 264906, 265009, 264596, 22278002
126	56701283 (251, 252)	Novel Protein sim. GBank gi|5102785|emb|CAB45200.1| - (AL079308) putative transcriptional regulator [Streptomyces coelicolor]			264511
127	20467267 (253, 254)			UNCLASSIFIED	264605
128	80248473 (255, 256)	Novel Protein sim. GBank gi|130120|sp|P23620|PHOB_PSEAE - PHOSPHATE REGULON TRANSCRIPTIONAL REGULATORY PROTEIN PHOB	Contains protein domain (PF00072) - Response regulator receiver domain	phosphatase	264907, 264909, 264910, 264600, 264601, 264603, 264605, 18108351, 264693, 264557
129	95290543 (257, 258)	Novel Protein sim. GBank gi|2506493|sp|P38036|YGCB_ECOLI - HYPOTHETICAL 100.5 KD PROTEIN IN IAP-CYSH INTERGENIC REGION	Contains protein domain (PF00270) - DEAD/DEAH box helicase	UNCLASSIFIED	35696423, 35695855, 264600, 264602, 264603, 264604, 264605, 264508, 264906, 264564, 264628, 264682, 264565, 264683
130	80085583 (259, 260)	Novel Protein sim. GBank gi|854065|emb|CAA58337| - (X83413) U88 [Human herpesvirus 6]			264634
131	94995022 (261, 262)	Novel Protein sim. GBank gi|1076038|pir||S54860 - ABC transporter PstC-2 chain - Mycobacterium tuberculosis	Contains protein domain (PF00528) - Binding-protein-dependent transport systems inner membrane component	transport	18108376, 264769, 29331826, 264689, 22278996, 265021, 264600, 264511, 264601, 264602, 264605, 264905, 264636

132	10887692 (263, 264)				264636
133	94630883 (265, 266)	Novel Protein sim. GBank gi|1877340|emb|CAB07068| - (Z92771) accA3 [Mycobacterium tuberculosis]	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPSase)	carboxylase	264905, 264689
134	79834660 (267, 268)	Novel Protein sim. GBank gi|4585838|emb|CAB40932.1| - (AL049630) putative NADH dehydrogenase [Streptomyces coelicolor]		dehydrogenase	264905, 264605, 265021
135	19885057 (269, 270)	Novel Protein sim. GBank gi|1460074|emb|CAB01049| - (Z77250) hypothetical protein Rv2566 [Mycobacterium tuberculosis]			264634
136	79846083 (271, 272)	Novel Protein sim. GBank gi|2125896|emb|CAA73511| - (Y13070) folylpolyglutamate synthase [Streptomyces coelicolor]		synthase	264508
137	79619770 (273, 274)				264683, 264685, 264686, 264691, 264692, 264693
138	79635971 (275, 276)	Novel Protein sim. GBank gi|5420387|emb|CAB46679.1| - (AJ243459) proteophosphoglycan [Leishmania major]		UNCLASSIFIED	18108374, 18108385, 33657109, 33657182, 265010, 22278998, 265006, 265007, 265008, 265009, 264693
139	86688076 (277, 278)	Novel Protein sim. GBank gi|5689912|emb|CAB52075.1| - (AL109732) putative mutase [Streptomyces coelicolor A3(2)]	Contains protein domain (PF01817) - Chorismate mutase	dehydrogenase	22278996, 265007, 264910, 60433356, 265010, 264602, 264605, 264766, 264688, 264769, 264693, 32833986, 18108374, 18108387
140	79825759 (279, 280)			UNCLASSIFIED	264908
141	20700094 (281, 282)				264600
142	80028104 (283, 284)	Novel Protein sim. GBank gi|3581916|emb|CAA20855| - (AL031545) mutS family DNA mismatch repair protein [Schizosaccharomyces pombe]		nuclease	264602, 265017
143	11072274 (285, 286)			UNCLASSIFIED	264600
144	95009102 (287, 288)	Novel Protein sim. GBank gi|3334127|sp|P97303|BAC2_MOUSE - TRANSCRIPTION REGULATOR PROTEIN BACH2 (BTB AND CNC HOMOLOG 2)			263978, 264600, 264910, 264632, 264508, 264563, 264564, 264591, 264556, 264908, 264629, 264639
145	80027058 (289, 290)	Novel Protein sim. GBank gi|3757569|emb|CAA21315| - (AL031863) 1-evidence=predicted by content; 1-method=genefinder;084; 1-method_score=66.31; 1-evidence_end [Drosophila melanogaster]		UNCLASSIFIED	22278996, 264602
146	13085662 (291, 292)	Novel Protein sim. GBank gi|140807|sp|P24536|YI21_BURCE - INSERTION ELEMENT IS402 HYPOTHETICAL 24 KD PROTEIN	Contains protein domain (PF01675) - Transposase		264687
147	94320366 (293, 294)	Novel Protein sim. GBank gi|2827608|emb|CAA16663| - (AL021646) uvrD2 [Mycobacterium tuberculosis]		helicase	264905, 264906, 264909, 264510, 265009, 60433356, 264600, 264601, 264604, 264605, 264687, 264769, 18108365, 65274791, 18108387
148	80248804 (295, 296)	Novel Protein sim. GBank gi|2916947|emb|CAA17585| - (AL021999) hypothetical protein Rv0986 [Mycobacterium tuberculosis]		transport	265009, 265010, 264600, 264602, 264603, 264604, 264605, 264693, 33657109, 264636

149	80249373 (297, 298)	Novel Protein sim. GBank gi|1723073|sp|Q11040|Y081_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.01	Contains protein domain (PF00005) - ABC transporter	transport	265010, 264600, 264601, 264603, 264604, 27486265, 264636
150	20294748 (299, 300)	Novel Protein sim. GBank gi|3724125|emb|CAA11905| - (AJ224340) maltosephosphorylase [Lactobacillus sanfrancisco]			264600
151	20726398 (301, 302)	Novel Protein sim. GBank gi|729312|sp|P07651|DEOB_ECOLI - PHOSPHOPENTOMUTASE (PHOSPHODEOXYRIBOMUTASE)	Contains protein domain (PF01676) - Metalloenzyme superfamily	UNCLASSIFIED	264602
152	95002877 (303, 304)	Novel Protein sim. GBank gi|2497952|sp|P55667|Y4TM_RHISN - HYPOTHETICAL HYDROLASE/PEPTIDASE Y4TM		peptidase	264602
153	80256665 (305, 306)	Novel Protein sim. GBank gi|3123021|sp|Q90508|VIT1_FUNHE - VITELLOGENIN I PRECURSOR (VTG I) (CONTAINS: LIPOVITELLIN 1 (LV1); PHOSVITIN (PV); LIPOVITELLIN 2 (LV2))		UNCLASSIFIED	264593
154	82305966 (307, 308)				264910, 264762, 264691, 264634
155	20429859 (309, 310)	Novel Protein sim. GBank gi|419697|pir||JN0443 - transcription initiation factor sigma homolog hrdB - Streptomyces aureofaciens	Contains protein domain (PF00140) - Sigma-70 factor	mapolymerase	264605
156	39564742 (311, 312)	Novel Protein sim. GBank gi|628710|pir||S41739 - hypothetical protein - Escherichia coli		UNCLASSIFIED	264565
157	10358887 (313, 314)	Novel Protein sim. GBank gi|3695013 (AF052586) - CtrA [Pseudomonas aeruginosa]	Contains protein domain (PF00142) - 4Fe-4S iron sulfur cluster binding proteins, NifH/frxC family	hydrolase	264691
158	79761936 (315, 316)	Novel Protein sim. GBank gi|1073072|pir||C55543 - cmaU protein - Pseudomonas syringae pv. syringae		UNCLASSIFIED	264905
159	78890376 (317, 318)			UNCLASSIFIED	265008
160	11075119 (319, 320)		Contains protein domain (PF00400) - WD domain, G-beta repeat		264605
161	80055007 (321, 322)	Novel Protein sim. GBank gi|1173023|sp|P46789|RL30_STRCO - 50S RIBOSOMAL PROTEIN L30	Contains protein domain (PF00327) - Ribosomal protein L30p/L7e	ribosomalprot	22278996, 264600, 264603, 35695917, 32833986, 35696423, 264636
162	80016371 (323, 324)	Novel Protein sim. GBank gi|5304869|emb|CAB46028.1| - (AL031685) dJ963K23.2 (novel protein) [Homo sapiens]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	interleukin	264112, 264532, 22279002
163	11692306 (325, 326)			UNCLASSIFIED	264639
164	80077902 (327, 328)			UNCLASSIFIED	264905, 264907, 264600
165	10856067 (329, 330)				264691
166	88095003 (331, 332)	Novel Protein sim. GBank gi|2661691|emb|CAA15795| - (AL009204) putative protease [Streptomyces coelicolor]		UNCLASSIFIED	264605, 264486
167	16395460 (333, 334)	Novel Protein sim. GBank gi|4416478|gb|AAD20378| - (AF125999) transposase [Mycobacterium avium]		UNCLASSIFIED	265010
168	80079362 (335, 336)	Novel Protein sim. GBank gi|76177|pir||QQECFT - hypothetical 38.8K protein (ftsI 5' region) - Escherichia coli			264600
169	80239581 (337, 338)				264556, 264557, 264558, 264559

170	79612364 (339, 340)				264906
171	95293073 (341, 342)	Novel Protein sim. GBank gi|140888|sp|P27847|YIGK_ECOLI - HYPOTHETICAL 15.4 KD PROTEIN IN RECQ-PLDB INTERGENIC REGION (F138)	Contains protein domain (PF01810) - LysE type translocator		264595, 264604
172	37797007 (343, 344)	Novel Protein sim. GBank gi|4210905|gb|AAD12048.1| - (AF045609) AglG [Sinorhizobium meliloti]	Contains protein domain (PF00528) - Binding-protein-dependent transport systems inner membrane component	transport	264769
173	57529660 (345, 346)	Novel Protein sim. GBank gi|132854|sp|P02387|RL2_ECOLI - 50S RIBOSOMAL PROTEIN L2	Contains protein domain (PF00181) - Ribosomal Proteins L2	ribosomalprot	264769
174	95293078 (347, 348)	Novel Protein sim. GBank gi|1881350|dbj|BAA19377| - (AB001488) PROBABLE TRANSPORT PROTEIN, SIMILAR TO ANTIBIOTIC TRANSPORT-ASSOCIATED PROTEIN ACTII IN STREPTOMYCES COELICOLOR. [Bacillus subtilis]		transport	264510, 264593, 264602, 264603, 264605, 264762, 264693
175	79756270 (349, 350)	Novel Protein sim. GBank gi|2072722|emb|CAB08326| - (Z95121) manA [Mycobacterium tuberculosis]		isomerase	264565
176	80066896 (351, 352)	Novel Protein sim. GBank gi|1055198 (U40187) - similar to PIR:A41724 chicken LD (limb deformity) gene product and to formin; also P-rich region similar to collagen [Caenorhabditis elegans]		UNCLASSIFIED	264907, 264910, 264681, 264558
177	86684852 (353, 354)	Novel Protein sim. GBank gi|2326738|emb|CAB10952| - (Z98268) hypothetical protein Rv1695 [Mycobacterium tuberculosis]	Contains protein domain (PF01513) - Domain of unknown function	UNCLASSIFIED	264768, 60424179, 264687, 264688, 264769, 29331826, 60432289, 18108376, 264689, 18108387, 32833986, 22278996, 265020, 264600, 264601, 264602, 264603, 264604, 264605, 264635, 264762, 264636, 264906, 264564, 264637, 264638, 264486, 60433356, 264766
178	79559526 (355, 356)	Novel Protein sim. GBank gi|1906596 (U81788) - kinesin-73 [Drosophila melanogaster]		struct	264693, 33657109, 264635
179	20263112 (357, 358)			UNCLASSIFIED	264563
180	80488958 (359, 360)	Novel Protein sim. GBank gi|1169367|sp|P45256|DNAB_HAEIN - REPLICATIVE DNA HELICASE		helicase	264769
181	79585369 (361, 362)	Novel Protein sim. GBank gi|3170615 (AF059485) - DOC4 [Mus musculus]		UNCLASSIFIED	21906767, 264635, 264639, 18108384
182	80577899 (363, 364)			UNCLASSIFIED	264259, 35696052, 56182435, 264511, 265018, 33657109, 264555, 264566
183	11614017 (365, 366)	Novel Protein sim. GBank gi|1076627|pir||S54172 - inorganic pyrophosphatase (EC 3.6.1.1) - common tobacco		UNCLASSIFIED	264690
184	10174167 (367, 368)	Novel Protein sim. GBank gi|4371280|gb|AAD18138| - (AC006260) hypothetical protein [Arabidopsis thaliana]		UNCLASSIFIED	264510

185	21660822 (369, 370)	Novel Protein sim. GBank gi|3006178|emb|CAA18398.1| - (AL022304) putative mma transport regulator [Schizosaccharomyces pombe]		UNCLASSIFIED	264604
186	80070329 (371, 372)	Novel Protein sim. GBank gi|2829802|sp|P94408|YCLF_BACSU - HYPOTHETICAL 53.3 KD PROTEIN IN SFP-GERKA INTERGENIC REGION		transport	264595
187	80186611 (373, 374)			UNCLASSIFIED	264369
188	20464942 (375, 376)	Novel Protein sim. GBank gi|3150260|emb|CAA19179| - (AL023634) cyclin [Schizosaccharomyces pombe]		kinase	264605
189	82338215 (377, 378)	Novel Protein sim. GBank gi|2145853|pir||S72938 - hflX protein - Mycobacterium leprae		UNCLASSIFIED	35696052, 264602, 264605, 264762, 264689, 35695917, 18108370, 18108372, 264638, 264565
190	80086821 (379, 380)	Novel Protein sim. GBank gi|1881244|dbj|BAA19271| - (AB001488) SIMILAR TO PYRUVATE OXIDASE AND ACETOLACTATE SYNTHASE. [Bacillus subtilis]	Contains protein domain (PF00205) - Thiamine pyrophosphate enzymes	synthase	264563
191	88095012 (381, 382)	Novel Protein sim. GBank gi|120226|sp|P28725|FKBP_STRCH - FK506-BINDING PROTEIN (PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (PPIASE) (ROTAMASE)	Contains protein domain (PF00254) - FKBP-type peptidyl-prolyl cis-trans isomerases	isomerase	264508, 264604, 264605, 264769, 264555
192	16333379 (383, 384)				264567
193	79910127 (385, 386)				264908, 264693
194	20464949 (387, 388)				264605
195	13518389 (389, 390)	Novel Protein sim. GBank gi|4980892|gb|AAD35474.1|AE00171 - (AE001718) ABC transporter, ATP-binding protein [Thermotoga maritima]		transport	264636
196	95005569 (391, 392)	Novel Protein sim. GBank gi|1705461|sp|P53656|BIOA_ERWHE - ADENOSYLMETHIONINE-8-AMINO-7-OXONONANOATE AMINOTRANSFERASE (7,8-DIAMINO-PELARGONIC ACID AMINOTRANSFERASE) (DAPA AMINOTRANSFERASE)	Contains protein domain (PF00202) - Aminotransferases class-III pyridoxal phosphate	gaba	264600, 264689, 264638
197	80248665 (393, 394)	Novel Protein sim. GBank gi|3122305|sp|Q27778|K6PF_SCHMA - 6-PHOSPHOFRUCTOKINASE (PHOSPHOFRUCTOKINASE) (PHOSPHOHEXOKINASE)	Contains protein domain (PF00365) - Phosphofructokinase	kinase	264602, 264682, 264692, 18108374
198	79163635 (395, 396)				264636
199	78890715 (397, 398)	Novel Protein sim. GBank gi|1781203|emb|CAB06110| - (Z83859) gnd [Mycobacterium tuberculosis]	Contains protein domain (PF00393) - 6-phosphogluconate dehydrogenases		265008
200	79413849 (399, 400)	Novel Protein sim. GBank gi|2642222 (AF030885) - telomere-associated recQ-like helicase [Ustilago maydis]		UNCLASSIFIED	264595, 264596
201	86945924 (401, 402)	Novel Protein sim. GBank gi|2894379|emb|CAA74911.1| - (Y14573) ring finger protein [Hordeum vulgare]		UNCLASSIFIED	29331826, 265007, 264512, 33657402, 264596, 265017, 18108351, 264682, 264683, 264767, 264629, 55810764, 264634, 264635, 56182323, 60432113, 22279000

202	79588046 (403, 404)	Novel Protein sim. GBank gi|231772|sp|P30598|CHS1_USTMA - CHITIN SYNTHASE 1 (CHITIN-UDP ACETYL-GLUCOSAMINYL TRANSFERASE 1)	Contains protein domain (PF01644) - Chitin synthase	synthase	264600
203	79843927 (405, 406)	Novel Protein sim. GBank gi|1504042|dbj|BAA13220| - (D86984) similar to yeast adenylate cyclase (S56776) [Homo sapiens]			22278995, 29331822, 29331825, 29331827, 264906, 21906754, 264683, 21906766, 21906769, 35696423, 264558
204	79855186 (407, 408)			UNCLASSIFIED	264909
205	10090583 (409, 410)	Novel Protein sim. GBank gi|2633808|emb|CAB13310| - (Z99111) similar to hypothetical proteins [Bacillus subtilis]		transport	264909
206	8758473 (411, 412)			UNCLASSIFIED	264604
207	20754522 (413, 414)	Novel Protein sim. GBank gi|2134381|pir||S60678 - polybromo 1 protein - chicken		UNCLASSIFIED	264556
208	20289261 (415, 416)				264605
209	80071069 (417, 418)	Novel Protein sim. GBank gi|2501040|sp|O05814|SYP_MYCTU - PROLYL-TRNA SYNTHETASE (PROLINE--TRNA LIGASE) (PRORS)			264605, 264689
210	80168800 (419, 420)				264905, 264907, 264909, 264766, 264687, 264691, 264629, 18108374, 264638
211	80034539 (421, 422)				263978
212	82442474 (423, 424)	Novel Protein sim. GBank gi|5031809|ref|NP_005536.1|pISLR - immunoglobulin superfamily containing leucine-rich repeat		UNCLASSIFIED	264508, 264905, 264906, 264907, 264908, 264600, 264762, 264534, 264632, 264634, 264635, 264639, 264486
213	80249562 (425, 426)	Novel Protein sim. GBank gi|3122359|sp|O33123|LEU2_MYCLE - 3-ISOPROPYLMALATE DEHYDRATASE LARGE SUBUNIT (ISOPROPYLMALATE ISOMERASE) (ALPHA-IPM ISOMERASE) (IPMI)	Contains protein domain (PF00330) - Aconitase family (aconitate hydratase)	isomerase	22278996, 264508, 264600, 264602, 264603, 264605, 33657023, 264565, 264486
214	80079381 (427, 428)	Novel Protein sim. GBank gi|116236|sp|P19421|CH60_COXBU - 60 KD CHAPERONIN (PROTEIN CPN60) (GROEL PROTEIN) (HEAT SHOCK PROTEIN B)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	264600, 264693
215	14973283 (429, 430)			UNCLASSIFIED	264629
216	80177716 (431, 432)	Novel Protein sim. GBank gi|3417297 (AC002310) - Unknown gene product [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	dna_rna_bind	264448
217	79603634 (433, 434)	Novel Protein sim. GBank gi|2506924|sp|P49754|VP41_HUMAN - VACUOLAR ASSEMBLY PROTEIN VPS41 HOMOLOG (S53)			264508
218	80258475 (435, 436)	Novel Protein sim. GBank gi|1173288|sp|P38106|RSEA_ECOLI - SIGMA-E FACTOR NEGATIVE REGULATORY PROTEIN		rnapolymerase	264594
219	20438797 (437, 438)	Novel Protein sim. GBank gi|1781097|emb|CAB06231| - (Z83864) gltB [Mycobacterium tuberculosis]		synthase	264604
220	13499572 (439, 440)	Novel Protein sim. GBank gi|2984703 (AF052427) - unknown [Trypanosoma cruzi]		nucleaseinhib	264689
221	11287498 (441, 442)	Novel Protein sim. GBank gi|4587313|dbj|BAA76709.1| - (AB025248) alpha-1,2-mannosidase [Bacillus sp. M-90]		UNCLASSIFIED	264555

222	79862802 (443, 444)	Novel Protein sim. GBank gi|1877268|emb|CAB07049| - (Z92770) hypothetical protein Rv0143c [Mycobacterium tuberculosis]		UNCLASSIFIED	264605, 264769, 35696423
223	83053869 (445, 446)			UNCLASSIFIED	264906, 264907, 264603
224	79557920 (447, 448)				264684, 264693
225	79559541 (449, 450)	Novel Protein sim. GBank gi|2274851|dbj|BAA21515| - (D64159) 3-7 gene product [Homo sapiens]		UNCLASSIFIED	264692
226	79172397 (451, 452)	Novel Protein sim. GBank gi|868245 (U29488) - C56C10.7 gene product [Caenorhabditis elegans]		UNCLASSIFIED	22278998, 264112, 33657023, 263981
227	81777196 (453, 454)			UNCLASSIFIED	35695917, 264636, 264907
228	79872285 (455, 456)				264768, 264907, 264908, 264692, 264593, 264639
229	79838266 (457, 458)				264906, 264910
230	11013209 (459, 460)			UNCLASSIFIED	264631
231	20622207 (461, 462)	Novel Protein sim. GBank gi|1835114|emb|CAA71733| - (Y10744) homoserine O-acetyltransferase [Leptospira meyeri]			264906, 264600, 264603, 264692
232	80055035 (463, 464)			UNCLASSIFIED	264600, 264603, 264605, 264687, 264769
233	80063054 (465, 466)	Novel Protein sim. GBank gi|2642340 (AF032970) - imidazolone propionate hydrolase [Pseudomonas putida]	Contains protein domain (PF00449) - Urease	hydrolase	264604
234	7523998 (467, 468)	Novel Protein sim. GBank gi|3510505 (AF030881) - pol polyprotein [Fugu rubripes]		UNCLASSIFIED	264369
235	80203671 (469, 470)			UNCLASSIFIED	264106
236	78940001 (471, 472)	Novel Protein sim. GBank gi|2104609|emb|CAB08805| - (Z95398) PckA [Mycobacterium leprae]		carboxylase	264905
237	11755273 (473, 474)				264681
238	79461401 (475, 476)			UNCLASSIFIED	264639
239	82435190 (477, 478)	Novel Protein sim. GBank gi|2495617|sp|Q57252|YDIJ_HAEIN - HYPOTHETICAL PROTEIN HI1163	Contains protein domain (PF00037) - 4Fe-4S ferredoxins and related iron-sulfur cluster binding domains.		264906, 265010, 264603, 264762, 264682, 264636, 264638, 264486
240	21635575 (479, 480)	Novel Protein sim. GBank gi|3183458|sp|P75796|YLIA_ECOLI - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YLIA		transport	264259, 264769
241	80377307 (481, 482)	Novel Protein sim. GBank gi|3875920|emb|CAB04111| - (Z81503) predicted using Genefinder; similar to collagen; cDNA EST EMBL:D65450 comes from this gene; cDNA EST EMBL:D68888 comes from this gene [Caenorhabditis elegans]		UNCLASSIFIED	264908, 264909, 264764, 264639
242	82148454 (483, 484)			UNCLASSIFIED	264489, 264907, 264908, 264511, 264760, 264764, 264692, 264635, 264637
243	79633207 (485, 486)			UNCLASSIFIED	264906
244	80248682 (487, 488)	Novel Protein sim. GBank gi|2624302|emb|CAA15575| - (AL008967) ald [Mycobacterium tuberculosis]		dehydrogenase	264600, 264602, 264605, 264769, 264689
245	79863543 (489, 490)	Novel Protein sim. GBank gi|2920625 (AF044499) - vgrE protein [Escherichia coli]		UNCLASSIFIED	264907, 264758
246	79162929 (491, 492)	Novel Protein sim. GBank gi|5420387|emb|CAB46679.1| - (AJ243459) proteophosphoglycan [Leishmania major]	Contains protein domain (PF01106) - NifU-like domain		264637, 18108381, 18108387, 264565

247	79873185 (493, 494)	Novel Protein sim. GBank gi|1839006|emb|CAB06648| - (Z85982) argB [Mycobacterium tuberculosis]		kinase	264909, 264691, 35696423, 18108387
248	80488983 (495, 496)	Novel Protein sim. GBank gi|1168574|sp|P42464|ATPB_CORGL - ATP SYNTHASE BETA CHAIN		synthase	35696286, 264907, 264511, 264602, 264768, 264688, 265021, 35695855, 18108385
249	79764645 (497, 498)			UNCLASSIFIED	264907, 264910, 265011, 264762, 264636
250	78619980 (499, 500)				21906768, 264692
251	84359489 (501, 502)			UNCLASSIFIED	52645156, 29331822, 29331824, 52644045, 265018, 21906765, 21906768, 265020, 27486261, 27486265, 35695763, 18108376, 264556, 264559, 264565
252	79737756 (503, 504)	Novel Protein sim. GBank gi|3327166|dbj|BAA31651| - (AB014576) KIAA0676 protein [Homo sapiens]			264685, 264687, 264632
253	20443124 (505, 506)	Novel Protein sim. GBank gi|3036880|emb|CAA18513| - (AL022374) putative ATP-dependent DNA helicase [Streptomyces coelicolor]		helicase	264604
254	80027421 (507, 508)	Novel Protein sim. GBank gi|3915488|sp|O34961|YJMB_BACSU - HYPOTHETICAL SYMPORTER IN COTT-RAPA INTERGENIC REGION		UNCLASSIFIED	264508, 264906, 264602, 264687, 265021, 264486
255	11398315 (509, 510)	Novel Protein sim. GBank gi|1665720|dbj|BAA04134| - (D17312) diarrheal toxin [Bacillus cereus]		UNCLASSIFIED	264593
256	80028158 (511, 512)	Novel Protein sim. GBank gi|465787|sp|P34422|YL31_CAEEL - HYPOTHETICAL 86.0 KD PROTEIN F44B9.1 IN CHROMOSOME III	Contains protein domain (PF00326) - Prolyl oligopeptidase family	peptidase	264602, 264692
257	20289282 (513, 514)	Novel Protein sim. GBank gi|1172039|sp|P42315|SCOA_BACSU - PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE SUBUNIT A (SUCCINYL COA:3-OXOACID COA-TRANSFERASE) (OXCT A)	Contains protein domain (PF01144) - Coenzyme A transferase	transferase	264605
258	20459464 (515, 516)	Novel Protein sim. GBank gi|3127836|emb|CAA18902| - (AL023496) hypothetical protein [Streptomyces coelicolor]		UNCLASSIFIED	264604
259	79910152 (517, 518)			collagen	264681, 264686, 264692
260	20379437 (519, 520)			UNCLASSIFIED	264692, 264556
261	20285883 (521, 522)	Novel Protein sim. GBank gi|123761|sp|P24221|HUTH_STRGR - HISTIDINE AMMONIA-LYASE (HISTIDASE)	Contains protein domain (PF00221) - Phenylalanine and histidine ammonia lyases	UNCLASSIFIED	264600
262	80189317 (523, 524)			UNCLASSIFIED	265017, 264369
263	88095045 (525, 526)	Novel Protein sim. GBank gi|3924708|emb|CAA84646| - (Z35597) Weak similarity with sea squirt nidogen precursor protein (blastp score 71); cDNA EST EMBL:T02069 comes from this gene; cDNA EST EMBL:D76135 comes from this gene; cDNA EST EMBL:D73147 comes from this gene; cDNA EST EMB...		UNCLASSIFIED	264488, 264905, 264906, 264907, 264908, 264909, 264512, 264910, 264758, 264596, 264604, 265019, 264605, 264760, 18108351, 264763, 264764, 264288, 264766, 264768, 264769, 264691, 264692, 264693, 264628, 264634, 264635, 264555, 264636, 264638, 264639
264	87370826 (527, 528)	Novel Protein sim. GBank gi|3043734|dbj|BAA25531| - (AB011177) KIAA0605 protein [Homo sapiens]	Contains protein domain (PF00047) - Immunoglobulin domain	protease	264259, 264908, 21906754, 265018, 265019, 265020

265	95355646 (529, 530)	Novel Protein sim. GBank gi|4589624|dbj|BAA76834.1| - (AB023207) KIAA0990 protein [Homo sapiens]		kinase .	264488, 35696286, 29331824, 56182181, 35696052, 264508, 264905, 264906, 264907, 66712502, 264908, 264909, 264511, 264512, 264910, 264592, 264595, 264758, 264596, 55811386, 264600, 265017, 264603, 264604, 264605, 264760, 18108351, 264762, 264681, 264764, 264288, 264766, 264768, 264769, 21906765, 21906767, 21906769, 265020, 264691, 33657023, 33657109, 33657182, 264628, 35696423, 35695855, 264630, 264631, 264632, 264634, 264635, 264636, 264555, 264638, 83373044, 56526486, 87168518, 264564, 264566, 264486
266	79588075 (531, 532)				264600
267	11362222 (533, 534)			UNCLASSIFIED	264828
268	79909566 (535, 536)			UNCLASSIFIED	264687, 264769, 264689
269	80025810 (537, 538)			UNCLASSIFIED	264602
270	84361144 (539, 540)	Novel Protein sim. GBank gi|4507367|ref|NP_003182.1|pTARS - threonyl-tRNA synthetase		UNCLASSIFIED	264693
271	79552301 (541, 542)			UNCLASSIFIED	264909, 264693
272	9674778 (543, 544)	Novel Protein sim. GBank gi|4980738|gb|AAD35331.1|AE00170 - (AE001707) glucose-1-phosphate adenylyltransferase [Thermotoga maritima]		synthase	264908
273	12840694 (545, 546)	Novel Protein sim. GBank gi|1168224|sp|P44569|5NTD_HAEIN - PROBABLE 5'-NUCLEOTIDASE PRECURSOR		UNCLASSIFIED	264688
274	39524246 (547, 548)				264564
275	82787041 (549, 550)	Novel Protein sim. GBank gi|3253159 (AF005355) - translation initiation factor eIF2C [Oryctolagus cuniculus]		UNCLASSIFIED	264907, 264908, 264909, 264766, 264768, 264691, 264632, 264636
276	86671073 (551, 552)	Novel Protein sim. GBank gi|134920|sp|P21997|SSGP_VOLCA - SULFATED SURFACE GLYCOPROTEIN 185 (SSG 185)			265008, 60432229
277	80079735 (553, 554)	Novel Protein sim. GBank gi|129021|sp|P20964|OBG_BACSU - SPO0B-ASSOCIATED GTP-BINDING PROTEIN		ribosomalprot	264600, 18108387
278	12866947 (555, 556)			UNCLASSIFIED	264689
279	95292719 (557, 558)	Novel Protein sim. GBank gi|79839|pir||S03812 - uvrB protein - Micrococcus luteus		nuclease	264508, 264604, 21906764, 264638, 264557, 264404
280	5603617 (559, 560)				264259
281	80249599 (561, 562)	Novel Protein sim. GBank gi|3123160|sp|Q18964|YLN2_CAEEL - HYPOTHETICAL 46.2 KD TRP-ASP REPEATS CONTAINING PROTEIN D2013.2 IN CHROMOSOME II			18108392, 264634, 264555, 264556, 264557, 264558
282	18598682 (563, 564)			UNCLASSIFIED	265019
283	20614211 (565, 566)			UNCLASSIFIED	264555

284	91212160 (567, 568)	Novel Protein sim. GBank gi|2429094 (U58632) - acetyl xylan esterase; AxeA [Thermotoga neapolitana]	Contains protein domain (PF00300) - Phosphoglycerate mutase family	UNCLASSIFIED	35696052, 29331828, 264508, 264905, 264600, 264602, 264605, 264682, 264764, 56181562, 21906764, 18108376, 264636, 264559, 18108387
285	8757940 (569, 570)			UNCLASSIFIED	264603
286	80503235 (571, 572)	Novel Protein sim. GBank gi|2072674|emb|CAB08305| - (Z95120) rhlE [Mycobacterium tuberculosis]	Contains protein domain (PF00270) - DEAD/DEAH box helicase	ATPase_associated	35696052, 264769, 264638
287	12745521 (573, 574)			UNCLASSIFIED	264689
288	20756502 (575, 576)	Novel Protein sim. GBank gi|765323|bbs|157676 - (S74439) silk fibroin heavy chain (C-terminal) [Bombyx mori=silkworms, Peptide Partial, 633 aa] [Bombyx mori]		collagen	264557
289	80043804 (577, 578)	Novel Protein sim. GBank gi|1870009|emb|CAB06860| - (Z92539) hypothetical protein Rv1019 [Mycobacterium tuberculosis]	Contains protein domain (PF00440) - Bacterial regulatory proteins, tetR family	ribosomalprot	264593, 264600
290	80430175 (579, 580)			UNCLASSIFIED	264768
291	20747431 (581, 582)	Novel Protein sim. GBank gi|2506664|sp|P40120|YDCG_ECOLI - 59.4 PROTEIN IN TRG-RIML INTERGENIC REGION PRECURSOR		UNCLASSIFIED	264601
292	80052555 (583, 584)	Novel Protein sim. GBank gi|625182 (L39015) - mitochondrial glutamyl-tRNA synthetase [Saccharomyces cerevisiae]		UNCLASSIFIED	264605
293	80062519 (585, 586)	Novel Protein sim. GBank gi|1718065|sp|P53528|UVRD_MYCLE - PUTATIVE DNA HELICASE II HOMOLOG		helicase	264909, 264605, 264687, 264689, 264692
294	79830303 (587, 588)	Novel Protein sim. GBank gi|117422|sp|P10040|CRB_DROME - CRUMBS PROTEIN PRECURSOR (95F)	Contains protein domain (PF00008) - EGF-like domain	oncogene	35696052, 264906, 265011, 264628, 55811576
295	79444180 (589, 590)	Novel Protein sim. GBank gi|1181619|dbj|BAA11565| - (D82364) a variant of TSC-22 [Gallus gallus]			52644507, 29331822, 264592, 265020, 264639
296	79607076 (591, 592)	Novel Protein sim. GBank gi|3649789|dbj|BAA33403| - (AB012226) SecA [Vibrio alginolyticus]		synthase	264508
297	79631297 (593, 594)	Novel Protein sim. GBank gi|5689967|emb|CAB52004.1| - (AL109663) putative membrane protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	264905, 264687, 264638
298	80418898 (595, 596)			UNCLASSIFIED	264905, 264691, 264639, 264766

299	95293298 (597, 598)	Novel Protein sim. GBank gi|220637|dbj|BAA01477| - (D10627) zinc finger protein [Mus musculus]	Contains protein domain (PF00096) - Zinc finger, C2H2 type		264488, 263994, 56994075, 22278997, 22278998, 22278999, 20281099, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 264905, 264906, 264907, 264908, 52644045, 264909, 264511, 265008, 264910, 264595, 264596, 264758, 33657084, 87168559, 265018, 265019, 264764, 264288, 264766, 264687, 56181562, 264769, 21906765, 21906768, 21906769, 33657023, 264692, 33657109, 27486281, 18108370, 264628, 264629, 55811576, 35695855, 264631, 264634, 264635, 264638, 264639, 83373044, 18108387, 87168518, 22279000, 22279002, 264565, 264566, 264567
300	20711340 (599, 600)			UNCLASSIFIED	264602
301	13511332 (601, 602)	Novel Protein sim. GBank gi|145922 (M20981) - iron dicitrate transport protein precursor [Escherichia coli]		transport	264687
302	9875260 (603, 604)	Novel Protein sim. GBank gi|1174661|sp|P44594|TGT_HAEIN - QUEUINE TRNA-RIBOSYLTRANSFERASE (TRNA-GUANINE TRANSGLYCOSYLASE) (GUANINE INSERTION ENZYME)			264908
303	79574895 (605, 606)				264689
304	20711344 (607, 608)	Novel Protein sim. GBank gi|67985|pir||HJNVAV - helicase (EC 3.6.1.-) - Autographa californica nuclear polyhedrosis virus		helicase	264602
305	80412520 (609, 610)	Novel Protein sim. GBank gi|728867|sp|P40602|APG_ARATH - ANTER-SPECIFIC PROLINE-RICH PROTEIN APG PRECURSOR			264763
306	8515876 (611, 612)	Novel Protein sim. GBank gi|1657554|gb|AAB18082.1| - (U73857) hypothetical protein [Escherichia coli]		UNCLASSIFIED	263978
307	80222901 (613, 614)			UNCLASSIFIED	265010, 21906768, 265020, 18108374, 263977
308	80064305 (615, 616)	Novel Protein sim. GBank gi|1710612|sp|Q10793|RNH2_MYCTU - PROBABLE RIBONUCLEASE HII (RNASE HII)	Contains protein domain (PF01351) - Ribonuclease HII	nuclease	264910, 264600, 264605, 264687, 264689, 264638, 18108387
309	80504136 (617, 618)	Novel Protein sim. GBank gi|5420387|emb|CAB46679.1| - (AJ243459) proteophosphoglycan [Leishmania major]			264769
310	80053616 (619, 620)				264603
311	11090659 (621, 622)	Novel Protein sim. GBank gi|1144522 (U34957) - phosphoribosylaminoimidazolesuccinocarboxamide synthase [Mycobacterium tuberculosis]		synthase	264602
312	80054347 (623, 624)			UNCLASSIFIED	264566
313	80046168 (625, 626)				264603, 264567

314	87645112 (627, 628)	Novel Protein sim. GBank gi|3661583 (AF092175) - ikaros [Danio rerio]	Contains protein domain (PF00320) - GATA zinc finger	dna_rna_bind	264259, 60432289, 29331828, 264905, 264906, 264908, 264909, 265008, 264910, 60432229, 33657402, 60433438, 33109954, 265011, 265017, 264603, 265018, 264288, 264766, 264692, 35695763, 264628, 264629, 264639, 60170394, 22279002, 264566
315	82356091 (629, 630)	Novel Protein sim. GBank gi|1652620|dbj|BAA17540| - (D90907) pyridine nucleotide transhydrogenase beta subunit [Synechocystis sp.]			264508, 264600, 264762, 264687, 264768, 52644229, 264769, 264689, 264635, 264636, 264638, 264486
316	79911071 (631, 632)			UNCLASSIFIED	264693
317	20466944 (633, 634)	Novel Protein sim. GBank gi|118244|sp|P24176|DAPE_ECOLI - SUCCINYL-DIAMINOPIMELATE DESUCCINYLASE (SDAP)		UNCLASSIFIED	264605
318	94141836 (635, 636)	Novel Protein sim. GBank gi|4680229|gb|AAD27583.1|AF11827 - (AF118274) DNb-5 [Homo sapiens]	Contains protein domain (PF00526) - Dictyostelium (slime mold) repeats	transport	264908, 264909, 264910, 264593, 264594, 264760, 264288, 264768, 264769, 21906769, 264691, 264693, 264628, 65274791, 264635, 264636, 264638, 83373044, 22279002, 264566
319	17289360 (637, 638)	Novel Protein sim. GBank gi|1149693|emb|CAA60220| - (X86499) rbsC [Clostridium perfringens]		transport	265018
320	13527675 (639, 640)	Novel Protein sim. GBank gi|2811033|sp|O05314|GLGC_MYCTU - GLUCOSE-1-PHOSPHATE ADENYLYLTRANSFERASE (ADP-GLUCOSE SYNTHASE) (ADP-GLUCOSE PYROPHOSPHORYLASE)		synthase	264687
321	94134387 (641, 642)	Novel Protein sim. GBank gi|1680716 (U68234) - all-trans-retinoic acid 4-hydroxylase [Danio rerio]		cyto450	264509, 264906, 264907, 264908, 265009, 264596, 264764, 264628, 264634, 264635, 264638, 264639, 83373044, 264567
322	66489053 (643, 644)	Novel Protein sim. GBank gi|1160355 (U33058) - UNC-89 [Caenorhabditis elegans]		UNCLASSIFIED	55811150, 264691, 60431528, 55810764
323	94653725 (645, 646)			UNCLASSIFIED	264488, 265009, 264593, 264628, 264635
324	79174383 (647, 648)				264687
325	79862691 (649, 650)			UNCLASSIFIED	264693
326	28774974 (651, 652)			UNCLASSIFIED	264288, 18108385
327	79776267 (653, 654)	Novel Protein sim. GBank gi|451544 (U04267) - proline-rich cell wall protein [Gossypium barbadense]			264488, 264905, 264509, 264910
328	80253202 (655, 656)			UNCLASSIFIED	264592
329	10173821 (657, 658)			UNCLASSIFIED	264510
330	86597767 (659, 660)	Novel Protein sim. GBank gi|4191358 (AF087825) - claudin-7 [Mus musculus]		UNCLASSIFIED	264259, 264908
331	79754888 (661, 662)	Novel Protein sim. GBank gi|80741|pir||S20912 - regulatory protein whiB - Streptomyces coelicolor		transcriptfactor	264910, 264687, 264689, 264636, 264567
332	80071440 (663, 664)	Novel Protein sim. GBank gi|114049|sp|P19480|AHPF_SALTY - ALKYL HYDROPEROXIDE REDUCTASE SUBUNIT F (ALKYL HYDROPEROXIDE REDUCTASE F52A PROTEIN)		reductase	35696423, 264636, 264638, 264565
333	13009555 (665, 666)				264687

334	80230771 (667, 668)	Novel Protein sim. GBank gi|322228|pir||S32227 - glutamate dehydrogenase (NADP+) (EC 1.4.1.4) - Corynebacterium glutamicum	Contains protein domain (PF00208) - Glutamate/Leucine/Phenylalanine/Valine dehydrogenase	dehydrogenase	264905, 264600, 264604, 264486
335	80057026 (669, 670)	Novel Protein sim. GBank gi|2193938|emb|CAB09602| - (Z96800) glpQ2 [Mycobacterium tuberculosis]		esterase	264907, 264603, 264693, 18108374, 264636, 18108387
336	80414319 (671, 672)			UNCLASSIFIED	265009, 264766, 264686
337	11090829 (673, 674)				264602
338	95413134 (675, 676)	Novel Protein sim. GBank gi|5454074|ref|NP_006303.1|pSMRT - silencing mediator for retinoid and thyroid hormone receptors	Contains protein domain (PF00249) - Myb-like DNA-binding domain	nucl_recpt	264569, 18108397, 22278998, 29331822, 20281099, 29331824, 56182181, 68714117, 29331825, 35696052, 29331828, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 265006, 265008, 264910, 265009, 264758, 55812038, 65274444, 265011, 87168559, 265017, 265018, 265019, 264760, 55811150, 264681, 264762, 18108351, 264682, 264764, 264766, 264685, 264686, 264768, 52644229, 264689, 55811957, 35695917, 264692, 264693, 264628, 18108370, 18108374, 55811576, 35696423, 35695855, 264635, 264555, 264636, 264556, 264637, 264557, 18108380, 264638, 264558, 264639, 18108381, 83373044, 18108385, 87168518, 60432113
339	11398513 (677, 678)	Novel Protein sim. GBank gi|4001713|dbj|BAA35087.1| - (AB015879) DnaK [Porphyromonas gingivalis]		eph	264593
340	80504149 (679, 680)	Novel Protein sim. GBank gi|2842699|sp|Q92353|UBPC_SCHPO - PUTATIVE UBIQUITIN CARBOXYL-TERMINAL HYDROLASE C6G9.08 (UBIQUITIN THIOLESTERASE) (UBIQUITIN-SPECIFIC PROCESSING PROTEASE) (DEUBIQUITINATING ENZYME)		ubiquitin	264905, 265019, 264769, 18108374
341	11075198 (681, 682)	Novel Protein sim. GBank gi|2688580 (AE001166) - conserved hypothetical protein [Borrelia burgdorferi]	Contains protein domain (PF00290) - Tryptophan synthase alpha chain	isomerase	264605
342	80054196 (683, 684)	Novel Protein sim. GBank gi|1684738|emb|CAA70601| - (Y09452) Yed j hypothetical protein [Pseudomonas syringae]			264603, 264604
343	20466792 (685, 686)				264605
344	80428870 (687, 688)	Novel Protein sim. GBank gi|2117275|emb|CAB09104| - (Z95618) hypothetical protein Rv0807 [Mycobacterium tuberculosis]		UNCLASSIFIED	264600, 264605, 264768, 18108370, 18108374, 35695855
345	80258853 (689, 690)	Novel Protein sim. GBank gi|3023317|sp|Q48935|APHA_MYCRA - ACETYLPOLYAMINE AMINOHYDROLASE		histone	264593
346	79831058 (691, 692)	Novel Protein sim. GBank gi|4239787|emb|CAA75437| - (Y15166) NADP-glutamate dehydrogenase [Pseudomonas aeruginosa]	Contains protein domain (PF00208) - Glutamate/Leucine/Phenylalanine/Valine dehydrogenase	dehydrogenase	264905

347	79158195 (693, 694)	Novel Protein sim. GBank gi|731675|sp|P38795|YHN4_YEAST - HYPOTHETICAL 80.7 KD PROTEIN IN ERG7-NMD2 INTERGENIC REGION		UNCLASSIFIED	265006, 265008, 265010, 265018, 263967, 263981
348	80020208 (695, 696)	Novel Protein sim. GBank gi|1073610|pir||S47672 - ugpB protein - Escherichia coli		transport	264602, 18108351, 18108387
349	17282112 (697, 698)				265007
350	80502370 (699, 700)	Novel Protein sim. GBank gi|3261599|emb|CAB00917| - (Z77137) hypothetical protein Rv1277 [Mycobacterium tuberculosis]		nuclease	265009, 264769, 264689, 18108370
351	80501805 (701, 702)	Novel Protein sim. GBank gi|2959367|emb|CAA17921| - (AL022117) hypothetical protein [Schizosaccharomyces pombe]		glycoprotein	264769, 264905, 264908
352	11611585 (703, 704)	Novel Protein sim. GBank gi|4416302|gb|AAD20307| - (AF105716) copia-type pol polyprotein [Zea mays]		protease	264595
353	80061653 (705, 706)	Novel Protein sim. GBank gi|1174887|sp|P42873|URE1_STAXY - UREASE ALPHA SUBUNIT (UREA AMIDOHYDROLASE)	Contains protein domain (PF00449) - Urease	UNCLASSIFIED	264604
354	56626130 (707, 708)			UNCLASSIFIED	264628
355	80046344 (709, 710)			UNCLASSIFIED	264909, 264595, 264683, 22279002
356	80043835 (711, 712)	Novel Protein sim. GBank gi|115157|sp|P16574|BVGA_BORPE - VIRULENCE FACTORS PUTATIVE POSITIVE TRANSCRIPTION REGULATOR BVGA	Contains protein domain (PF00072) - Response regulator receiver domain	transcriptfactor	264909, 264591, 264592
357	80070566 (713, 714)	Novel Protein sim. GBank gi|497637 (J03939) - cytochrome oxidase d subunit I [Escherichia coli]		oxidase	264605
358	37032756 (715, 716)	Novel Protein sim. GBank gi|2290990 (AF006000) - Brg1 [Bordetella pertussis]		UNCLASSIFIED	264768
359	80501488 (717, 718)			UNCLASSIFIED	264604, 264769
360	80026748 (719, 720)			UNCLASSIFIED	264594
361	80584075 (721, 722)	Novel Protein sim. GBank gi|3510639 (AF049344) - UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase T5 [Rattus norvegicus]		transferase	22278996, 264259, 29331822, 29331824, 264605, 55811957, 265022
362	13089485 (723, 724)	Novel Protein sim. GBank gi|113764|sp|P25718|AMY1_ECOLI - ALPHA-AMYLASE PRECURSOR (1,4-ALPHA-D-GLUCAN GLUCANOHYDROLASE)		amylase	264688
363	79750145 (725, 726)				264566
364	82443593 (727, 728)	Novel Protein sim. GBank gi|2829816|sp|P95171|NUOK_MYCTU - NADH DEHYDROGENASE I CHAIN K (NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 11) (NUO11)	Contains protein domain (PF00420) - NADH-ubiquinone/plastoquinone oxidoreductase chain 4L	dehydrogenase -	264769, 264602, 264604, 264508, 264762, 264638, 264486

365	88040288 (729, 730)	Novel Protein sim. GBank gi|4929268|gb|AAD33924.1| - (AF144237) LOMP protein [Homo sapiens]	Contains protein domain (PF00412) - LIM domain containing proteins		264488, 21906766, 21906767, 55811576, 21906769, 29148629, 22278995, 22278996, 265020, 265022, 264634, 264691, 264593, 33657023, 33657402, 264693, 264639, 264594, 29331824, 264758, 18108385, 29331827, 87168559, 265018, 22279000, 265019, 264482, 264761, 264681, 18108351
366	81821838 (731, 732)				265017, 264757
367	95357471 (733, 734)	Novel Protein sim. GBank gi|4503843|ref|NP_003908.1|pG2AD - UNKNOWN	Contains protein domain (PF01602) - Adaptin N terminal region	glycoprotein	60424179, 65274572, 56182575, 22278994, 56994075, 22278998, 264259, 29331822, 29331824, 56182181, 60424269, 66714117, 29331825, 60432289, 29331826, 29331827, 29331828, 264905, 264828, 56182435, 265006, 264512, 265008, 264591, 55812038, 55811386, 265010, 87168559, 265017, 265018, 264604, 265019, 55811150, 264448, 264369, 264288, 264686, 264768, 56181562, 21906768, 21906769, 55811957, 35695917, 265022, 60170615, 33657023, 65274620, 18108365, 263967, 33657109, 33657349, 35695763, 264628, 18108376, 55811576, 65274791, 35695855, 56182323, 83373044, 60432113, 264563, 264564, 264567
368	79607265 (735, 736)				264509
369	95292917 (737, 738)	Novel Protein sim. GBank gi|3913029|sp|P94967|ALR_MYCSM - ALANINE RACEMASE		UNCLASSIFIED	264508, 264604, 264605, 264636
370	88090966 (739, 740)	Novel Protein sim. GBank gi|3249559 (AF018261) - EH domain binding protein Epsin [Rattus norvegicus]			264905, 264592, 264605, 264766, 264691
371	95292599 (741, 742)	Novel Protein sim. GBank gi|2995299|emb|CAA18328| - (AL022268) putative tRNA delta(2)-isopentenylpyrophosphate transferase [Streptomyces coelicolor]	Contains protein domain (PF01715) - IPP transferase	transferase	264905, 264906, 264510, 264600, 264601, 264602, 264603, 265018, 264604, 264605, 265021, 264692, 264636, 264564
372	80021107 (743, 744)	Novel Protein sim. GBank gi|2506393|sp|P31576|FIXX_ECOLI - FERREDOXIN LIKE PROTEIN			264564
373	79863766 (745, 746)			UNCLASSIFIED	264909
374	79847568 (747, 748)	Novel Protein sim. GBank gi|3341640|emb|CAA13164| - (AJ231122) z61f [Vibrio cholerae]		UNCLASSIFIED	264905, 264906
375	91230181 (749, 750)	Novel Protein sim. GBank gi|5456934|gb|AAD43716.1| - (AF152322) protocadherin gamma A2 [Homo sapiens]		cadherin	65274572, 264259, 29331826, 56182435, 60433356, 60433438, 264757, 55812038, 264758, 55811957, 264690, 33657023
376	80505214 (751, 752)	Novel Protein sim. GBank gi|1805408|dbj|BAA08970| - (D50453) homologues to nitrile hydratase region 3'-hypothetical protein P47K of P. chlororaphis [Bacillus subtilis]		UNCLASSIFIED	264769
377	10339083 (753, 754)				264906

378	80056153 (755, 756)			UNCLASSIFIED	265008, 264555
379	80503437 (757, 758)	Novel Protein sim. GBank gi|1076013|pir||A49930 - carB protein homolog - Mycobacterium bovis (strain BCG) (fragment)	Contains protein domain (PF00289) - Carbamoyl-phosphate synthase (CPSase)	synthase	264769
380	80060937 (759, 760)	Novel Protein sim. GBank gi|216556|dbj|BAA02174| - (D12651) glucose dehydrogenase [Escherichia coli]	Contains protein domain (PF01011) - PQQ enzyme repeat	dehydrogenase	264604
381	11769027 (761, 762)			UNCLASSIFIED	264684
382	80054377 (763, 764)				264592
383	83259025 (765, 766)	Novel Protein sim. GBank gi|3327136|dbj|BAA31636| - (AB014561) KIAA0661 protein [Homo sapiens]			264595, 265017, 265021, 264638, 87168518, 22279002
384	95314255 (767, 768)			UNCLASSIFIED	264259, 29331822, 60432289, 29331827, 264288, 264766, 263987, 85274791, 35695855, 263981, 83373044, 264567
385	10237679 (769, 770)				264692
386	79633434 (771, 772)	Novel Protein sim. GBank gi|1073456|pir||S47810 - probable alcohol dehydrogenase (EC 1.1.1.1) - Escherichia coli	Contains protein domain (PF00465) - Iron-containing alcohol dehydrogenases	dehydrogenase	264906
387	17960637 (773, 774)	Novel Protein sim. GBank gi|1460074|emb|CAB01049| - (Z77250) hypothetical protein Rv2566 [Mycobacterium tuberculosis]	Contains protein domain (PF01841) - Transglutaminase-like superfamily	UNCLASSIFIED	264760
388	87741376 (775, 776)	Novel Protein sim. GBank gi|4240169|dbj|BAA74863.1| - (AB020647) KIAA0840 protein [Homo sapiens]	Contains protein domain (PF00646) - F-box domain.	homeobox	35696286, 264905, 66712502, 60432229, 264593, 60433356, 264686, 264688, 21906765, 264691, 22279000, 264482
389	79316971 (777, 778)			UNCLASSIFIED	18108394, 22278996, 264630, 264556, 22279002
390	80079949 (779, 780)			UNCLASSIFIED	264600
391	7657302 (781, 782)	Novel Protein sim. GBank gi|854065|emb|CAA58337| - (X83413) U88 [Human herpesvirus 6]			264482
392	79796056 (783, 784)			UNCLASSIFIED	264908
393	33206031 (785, 786)	Novel Protein sim. GBank gi|3378523|emb|CAA08867| - (AJ009832) cyclomaltodextrinase glucanotransferase [Thermotoga neapolitana]		synthase	264602, 21906764
394	10104463 (787, 788)				264693
395	80229010 (789, 790)			UNCLASSIFIED	264508, 264563
396	20436224 (791, 792)	Novel Protein sim. GBank gi|2677780 (U70327) - unknown [Paretroplus polyactis]	Contains protein domain (PF00047) - Immunoglobulin domain	struct	264556
397	80417014 (793, 794)	Novel Protein sim. GBank gi|4507909|ref|NP_000368.1|pWAS| - Wiskott-Aldrich syndrome (ecezema-thrombocytopenia)			265007, 265009, 264508, 264556, 264629, 264766
398	91230517 (795, 796)	Novel Protein sim. GBank gi|1518458 (U45998) - mitochondrial solute carrier [Onchocerca volvulus]	Contains protein domain (PF00153) - Mitochondrial carrier proteins	transport	18108398, 22278995, 22278996, 56994075, 22278999, 264259, 29331824, 29331826, 264905, 264908, 265007, 265008, 265009, 21906754, 33657084, 265017, 264448, 264288, 264766, 21906765, 21906766, 21906767, 265020, 265021, 33657023, 33657109, 264628, 35696423, 35695855, 264952, 18108380, 264567, 18108391

399	80055278 (797, 798)	Novel Protein sim. GBank gi|3358091|dbj|BAA31995| - (AB015974) glycerol kinase [Pseudomonas tolaasii]	Contains protein domain (PF00370) - FGGY family of carbohydrate kinases	kinase	264592, 264595
400	94117490 (799, 800)	Novel Protein sim. GBank gi|728835|sp|P39192|ALU5_HUMAN - !!!! ALU SUBFAMILY SC WARNING ENTRY !!!!	Contains protein domain (PF00560) - Leucine Rich Repeat	cadherin	18108394, 56182575, 22278995, 22278997, 22278999, 264259, 29331824, 265006, 265007, 265009, 60432229, 33657402, 21906754, 265010, 265017, 265018, 265019, 18108351, 18108357, 21906765, 265021, 265022, 264691, 264692, 33657023, 18108370, 65274791, 264634, 264636, 60170394, 56182323
401	11397491 (801, 802)	Novel Protein sim. GBank gi|4928292|gb|AAD33527.1|AF13211 - (AF132117) FhuA [Staphylococcus aureus]		transport	264594
402	95420294 (803, 804)	Novel Protein sim. GBank gi|5689487|dbj|BAA83027.1| - (AB028998) KIAA1075 protein [Homo sapiens]	Contains protein domain (PF00017) - Src homology domain 2	phosphatase	65274572, 56182575, 35696286, 22278996, 22278998, 264093, 264259, 29331822, 29331824, 29331825, 29331826, 60432289, 29331827, 29331828, 264906, 264907, 264909, 265006, 264511, 265007, 265008, 264910, 264591, 33657402, 60433356, 60433438, 264596, 21906754, 52644296, 265010, 265011, 87168559, 265017, 265018, 265019, 264681, 18108351, 264682, 264448, 264288, 264684, 264766, 264767, 264686, 21906765, 21906766, 21906767, 21906768, 21906769, 55811957, 265020, 265021, 265022, 264690, 264693, 65274620, 35695763, 264628, 18108370, 264629, 18108379, 35696423, 55811576, 264635, 264636, 264557, 264639, 18108385, 22279002, 264563, 264564, 264565, 264566
403	80439913 (805, 806)			UNCLASSIFIED	264768, 264632, 264639, 264563
404	11809865 (807, 808)				264682
405	79471280 (809, 810)	Novel Protein sim. GBank gi|2661649|emb|CAA15755| - (AL009198) dnaE2 [Mycobacterium tuberculosis]		polymerase	265009, 264682
406	79634172 (811, 812)		Contains protein domain (PF00159) - Pancreatic hormone peptides	UNCLASSIFIED	18108357, 264693
407	80478229 (813, 814)			UNCLASSIFIED	264769
408	80079956 (815, 816)			UNCLASSIFIED	264600
409	5640527 (817, 818)	Novel Protein sim. GBank gi|3047117 (AF058919) - similar to ATP-dependent RNA helicases [Arabidopsis thaliana]		helicase	264259

410	95357496 (819, 820)	Novel Protein sim. GBank gi|475016|dbj|BAA06184| - (D29801) Unknown [Mus musculus]		UNCLASSIFIED	264489, 52646365, 52646842, 56181686, 35696286, 52645080, 29331822, 29331824, 56182181, 29331825, 60424269, 35696052, 33656970, 264508, 264509, 264905, 264906, 264907, 264908, 52644045, 264909, 264510, 265007, 264512, 265008, 264910, 33657402, 264758, 52646317, 55811388, 265010, 265011, 265017, 264604, 265018, 55811150, 264762, 264764, 264766, 264687, 264768, 264769, 52644229, 21906766, 265020, 265021, 264534, 52644150, 264692, 33657023, 65274620, 33657109, 33657182, 27486261, 35695763, 264628, 264629, 60431528, 18108376, 263978, 35696423, 35695855, 264632, 264634, 264635, 264637, 264638, 264558, 264639, 56182323, 264559, 60432113, 22279002, 264563, 264565, 264486
411	80501670 (821, 822)			UNCLASSIFIED	264769
412	80241662 (823, 824)				264907, 264910, 263973, 22279002
413	11076446 (825, 826)	Novel Protein sim. GBank gi|3261784|emb|CAB08997| - (Z95558) htpX [Mycobacterium tuberculosis]		eph	264605
414	82050554 (827, 828)	Novel Protein sim. GBank gi|129036|sp|P20707|ODO1_AZOVI - 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (ALPHA-KETOGLUTARATE DEHYDROGENASE)		dehydrogenase	18108374, 264760, 264769, 264602, 264638, 264603, 264909, 264605
415	84453144 (829, 830)	Novel Protein sim. GBank gi|4868350|gb|AAD31273.1|AF13202 - (AF132025) rhophilin [Drosophila melanogaster]		UNCLASSIFIED	264908, 87168518
416	80402775 (831, 832)	Novel Protein sim. GBank gi|2555172 (AF025543) - ArcC; carbamate kinase [Rhizobium etli]		kinase	264488, 264600, 264602, 264764, 264636
417	20153797 (833, 834)	Novel Protein sim. GBank gi|1709171|sp|P52311|MTX2_XANOR - MODIFICATION METHYLASE XORII (CYTOSINE-SPECIFIC METHYLTRANSFERASE XORII) (M.XORII)	Contains protein domain (PF00145) - C-5 cytosine-specific DNA methylase		264605
418	94125841 (835, 836)			UNCLASSIFIED	264689, 264693
419	95314273 (837, 838)			collagen	264908, 264910, 264764, 264639
420	37036349 (839, 840)	Novel Protein sim. GBank gi|3261659|emb|CAB03751| - (Z81368) hypothetical protein Rv2419c [Mycobacterium tuberculosis]	Contains protein domain (PF00300) - Phosphoglycerate mutase family	phosphatase	264769
421	95292942 (841, 842)	Novel Protein sim. GBank gi|2916942|emb|CAA17580| - (AL021999) hypothetical protein Rv0981 [Mycobacterium tuberculosis]	Contains protein domain (PF00072) - Response regulator receiver domain	phosphatase	264906, 264600, 264601, 264603, 264604, 264760, 264769
422	79471293 (843, 844)	Novel Protein sim. GBank gi|231752|sp|Q00767|CH61_STRAL - 60 KD CHAPERONIN 1 (PROTEIN CPN60 1) (GROEL PROTEIN 1) (HSP58)	Contains protein domain (PF00118) - TCP-1/cpn60 chaperonin family	eph	22278996, 264682, 18108376, 18108387
423	79604948 (845, 846)			UNCLASSIFIED	264509

424	78966557 (847, 848)	Novel Protein sim. GBank gi|4826814|ref|NP_004977.1|pKTN1 - kinectin 1 (kinesin receptor)		struct	265019
425	80431450 (849, 850)	Novel Protein sim. GBank gi|1703701|bbs|178462 - KRP5=kinesin-related protein [rats, testes, Peptide Partial, 167 aa]	Contains protein domain (PF00225) - Kinesin motor domain	struct	264909, 265007, 55811386, 264768, 55810764
426	80064522 (851, 852)				264605, 264559
427	80057232 (853, 854)	Novel Protein sim. GBank gi|231829|sp|P29929|COBN_PSEDE - COBN PROTEIN		UNCLASSIFIED	264603, 264636
428	79487798 (855, 856)			UNCLASSIFIED	264683
429	80091252 (857, 858)	Novel Protein sim. GBank gi|81286|pir||S22697 - extensin - Volvox carteri (fragment)		UNCLASSIFIED	35696423, 35695763, 35695855, 265017, 264564, 264762
430	80504192 (859, 860)	Novel Protein sim. GBank gi|1806154|emb|CAB06451| - (Z84395) hypothetical protein Rv0688 [Mycobacterium tuberculosis]		reductase	264508, 264905, 264509, 264908, 264909, 265008, 264600, 264687, 264769, 264689, 264636, 264638, 18108385, 264486
431	20624249 (861, 862)				264566
432	16525372 (863, 864)				265020
433	81494303 (865, 866)	Novel Protein sim. GBank gi|3123552|emb|CAA18609| - (AL022578) dJ393P12.2 (hypothetical Proline-rich protein KIAA0269 LIKE) [Homo sapiens]		UNCLASSIFIED	264907, 264908, 264909, 264910, 264592, 264595, 264758, 264604, 264760, 264762, 264763, 264636, 264637, 22279002
434	94326323 (867, 868)	Novel Protein sim. GBank gi|2495272|sp|Q99626|CDX2_HUMAN - HOMEOBOX PROTEIN CDX-2 (CAUDAL-TYPE HOMEOBOX PROTEIN 2) (CDX-3)	Contains protein domain (PF00169) - PH domain	UNCLASSIFIED	55812038, 56182181, 56181562, 29331828, 35696052, 55810764, 55811576, 65274791, 35695855, 60432113, 55811150, 264636, 264766
435	80502738 (869, 870)	Novel Protein sim. GBank gi|114105|sp|P08532|ARAH_ECOLI - L-ARABINOSE TRANSPORT SYSTEM PERMEASE PROTEIN ARAH		transport	264595, 264769
436	41085953 (871, 872)			UNCLASSIFIED	265020, 22279002
437	11399291 (873, 874)			UNCLASSIFIED	264593
438	11773835 (875, 876)			UNCLASSIFIED	264686
439	80019495 (877, 878)	Novel Protein sim. GBank gi|3242702 (AC003040) - hypothetical protein [Arabidopsis thaliana]			264905, 264600, 264602, 264604
440	79841062 (879, 880)	Novel Protein sim. GBank gi|2291232|gb|AAB65351.1| - (AF016427) Contains similarity to Pfam domain: PF00004 (AAA), Score=268.1, E-value=3.7e-77, N=1 [Caenorhabditis elegans]	Contains protein domain (PF00004) - ATPases associated with various cellular activities (AAA)	ATPase_associated	35696052, 264905, 264908, 264909, 265011, 35696423
441	20396935 (881, 882)	Novel Protein sim. GBank gi|5639946|gb|AAD45904.1|AF16132 - (AF161328) histidine kinase CstS [Corynebacterium diphtheriae]			264605
442	85281058 (883, 884)	Novel Protein sim. GBank gi|1184790 (U46068) - von Ebner minor salivary gland protein [Mus musculus]		UNCLASSIFIED	29331830, 264909
443	82456427 (885, 886)	Novel Protein sim. GBank gi|5689893|emb|CAB52056.1| - (AL109732) putative ATP-binding RNA helicase [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	35696052, 264508, 264906, 264512, 264604, 264762, 264769, 264689, 264636
444	11395897 (887, 888)	Novel Protein sim. GBank gi|1783249|dbj|BAA11726| - (D83026) homologous to citrate-sodium symport (citrate transporters); hypothetical [Bacillus subtilis]		UNCLASSIFIED	264591

445	79552709 (889, 890)			UNCLASSIFIED	264693
446	79810937 (891, 892)	Novel Protein sim. GBank gi|5531272|emb|CAB50897.1| - (AJ243800) WSC4 homologue [Kluyveromyces lactis]			264509
447	80438888 (893, 894)	Novel Protein sim. GBank gi|538413 (L36315) - zinc finger protein [Mus musculus]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	264768, 55811576
448	80238110 (895, 896)	Novel Protein sim. GBank gi|1542914|emb|CAB02185| - (Z80108) fmt [Mycobacterium tuberculosis]	Contains protein domain (PF00551) - Formyl transferase	dehydrogenase	264508, 264600, 264603, 264605, 264682, 264769, 18108362, 264634, 18108387
449	20460634 (897, 898)	Novel Protein sim. GBank gi|118794|sp|P10443|DP3A_ECOLI - DNA POLYMERASE III, ALPHA CHAIN		polymerase	264605, 264559
450	94631210 (899, 900)	Novel Protein sim. GBank gi|4589506|dbj|BAA76775.1| - (AB023148) KIAA0931 protein [Homo sapiens]	Contains protein domain (PF00481) - Protein phosphatase 2C	phosphatase	65274572, 22278998, 29331824, 29331826, 264906, 264910, 264592, 52646317, 265017, 21906767, 55811957, 56526486, 22279002
451	21433609 (901, 902)			UNCLASSIFIED	264486
452	10267276 (903, 904)				264692
453	52560096 (905, 906)	Novel Protein sim. GBank gi|2650614 (AE001104) - conserved hypothetical protein [Archaeoglobus fulgidus]		UNCLASSIFIED	264907, 264600
454	39523922 (907, 908)	Novel Protein sim. GBank gi|2493000|sp|Q09450|SCOT_CAEEL - PROBABLE SUCCINYL-COA:3-KETOACID-COENZYME A TRANSFERASE PRECURSOR (3-OXOACID COA-TRANSFERASE)		transferase	264603
455	13089692 (909, 910)			UNCLASSIFIED	264687
456	79563081 (911, 912)			UNCLASSIFIED	264691
457	79831273 (913, 914)	Novel Protein sim. GBank gi|4468699|emb|CAB38153.1| - (AL035591) putative integral membrane export protein [Streptomyces coelicolor]			264905
458	79581227 (915, 916)	Novel Protein sim. GBank gi|3411053 (AF034863) - synaptic scaffolding molecule [Rattus norvegicus]	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF).	kinase	55812038, 265010, 265018, 264681
459	80567359 (917, 918)	Novel Protein sim. GBank gi|4506075|ref|NP_002733.1|pPRKC - protein kinase C, mu	Contains protein domain (PF00130) - Phorbol esters/diacylglycerol binding domain (C1 domain)	kinase	22278997, 264259, 29331826, 265018, 264448, 264369, 21906765, 35696423
460	79245890 (919, 920)	Novel Protein sim. GBank gi|113158|sp|P25516|ACO1_ECOLI - ACONITATE HYDRATASE 1 (CITRATE HYDRO-LYASE 1) (ACONITASE 1)		UNCLASSIFIED	264906
461	95287618 (921, 922)	Novel Protein sim. GBank gi|1168574|sp|P42464|ATPB_CORGL - ATP SYNTHASE BETA CHAIN		synthase	264602, 264605, 264768, 264769, 265021, 33657023, 264559

462	79606589 (923, 924)	Novel Protein sim. GBank gi|1346891|sp|P45597|PTF1_XANCP - MULTIPHOSPHORYL TRANSFER PROTEIN (MTP) (CONTAINS: PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I); PHOSPHOCARRIER PROTEIN HPR (PROTEIN H); PTS SYSTEM, FRUCTOSE-SPECIFIC IIA COMPONENT ...	Contains protein domain (PF00391) - PEP-utilizing enzymes	UNCLASSIFIED	264907
463	79796417 (925, 926)	Novel Protein sim. GBank gi|854065|emb|CAA58337| - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264905, 264906, 264908, 264909, 264910, 264591, 264595, 265011, 264632, 264635, 264636, 264637, 264638, 264639
464	82340151 (927, 928)			UNCLASSIFIED	264634
465	83005730 (929, 930)	Novel Protein sim. GBank gi|5689776|emb|CAB52137.1| - (AJ242832) calpain [Homo sapiens]	Contains protein domain (PF00648) - Calpain family cysteine protease	cathepsin	265017, 21906764, 265020
466	20460645 (931, 932)	Novel Protein sim. GBank gi|1806175|emb|CAB06470| - (Z84395) rpsC [Mycobacterium tuberculosis]	Contains protein domain (PF00417) - Ribosomal protein S3, N-terminal domain.	ribosomalprot	264605, 264559
467	80409035 (933, 934)	Novel Protein sim. GBank gi|548705|sp|P36949|RBSB_BACSU - D-RIBOSE-BINDING PROTEIN PRECURSOR		UNCLASSIFIED	264764
468	52562208 (935, 936)			UNCLASSIFIED	264692
469	19520527 (937, 938)	Novel Protein sim. GBank gi|2114024|emb|CAB08957| - (Z95558) grcC1 [Mycobacterium tuberculosis]		UNCLASSIFIED	264488
470	80502756 (939, 940)	Novel Protein sim. GBank gi|2909459|emb|CAA17347| - (AL021929) cobQ [Mycobacterium tuberculosis]		synthase	264602, 264769
471	17937351 (941, 942)	Novel Protein sim. GBank gi|114921|sp|P17447|BETT_ECOLI - HIGH-AFFINITY CHOLINE TRANSPORT PROTEIN		transport	265019
472	80047458 (943, 944)				264596, 264685, 264557
473	20558793 (945, 946)	Novel Protein sim. GBank gi|862343 (L10908) - Gcap1 gene product [Mus musculus]		UNCLASSIFIED	264369
474	80593365 (947, 948)			UNCLASSIFIED	22278997, 264692, 264288
475	82454665 (949, 950)			UNCLASSIFIED	264907, 264908, 264511, 265009, 264762, 264448, 264636, 264638
476	94143857 (951, 952)	Novel Protein sim. GBank gi|5453656|ref|NP_006329.1|pGAC1 - glioma amplified on chromosome 1 protein (leucine-rich)	Contains protein domain (PF00560) - Leucine Rich Repeat	glycoprotein	65274572, 60432049, 264259, 264508, 52644045, 55812038, 264758, 265011, 264288, 264686, 52644229, 65274791, 264638, 264566
477	79175833 (953, 954)			UNCLASSIFIED	264636
478	79633483 (955, 956)			UNCLASSIFIED	264690, 264693
479	80189746 (957, 958)			collagen	264686, 35695855, 265008, 264631, 264910, 264632, 264638, 265018, 264369, 264909
480	79390729 (959, 960)	Novel Protein sim. GBank gi|1127551 (U18939) - orf2 [Battrachocottus baikalensis]		mapolymerase	264369
481	79624578 (961, 962)			UNCLASSIFIED	264693
482	83050611 (963, 964)	Novel Protein sim. GBank gi|4063042 (AF068065) - GP900; mucin-like glycoprotein [Cryptosporidium parvum]		UNCLASSIFIED	264909, 264686, 264768, 264693, 55811576, 56182323, 18108385

483	20293306 (965, 966)	Novel Protein sim. GBank gi|2104303|emb|CAB08632| - (Z95387) hypothetical protein Rv2610c [Mycobacterium tuberculosis]	Contains protein domain (PF00534) - Glycosyl transferases group 1		264600
484	11618046 (967, 968)	Novel Protein sim. GBank gi|3450883 (AF083334) - fibroin [Antheraea pernyi]		UNCLASSIFIED	264594
485	80191234 (969, 970)			UNCLASSIFIED	264369, 21906765, 22279000, 22279002
486	80059042 (971, 972)	Novel Protein sim. GBank gi|5042272|emb|CAB44526.1| - (AL078618) nuoF, NADH dehydrogenase subunit [Streptomyces coelicolor]		dehydrogenase	264604
487	11813339 (973, 974)				264638
488	91222383 (975, 976)	Novel Protein sim. GBank gi|5724778|gb|AAC53522.2| - (AF012273) rho-type GTPase-activating protein rhoGAPX-1 [Mus musculus]	Contains protein domain (PF00620) - RhoGAP domain		264686, 66714117, 264768, 18108385, 55811576, 265006, 265008, 265009, 265019, 22279002, 264259, 18108370, 264907, 264764, 56182323, 264288, 264693
489	10867710 (977, 978)	Novel Protein sim. GBank gi|3882223|dbj|BAA34471.1| - (AB018294) KIAA0751 protein [Homo sapiens]		kinase	264639
490	95361124 (979, 980)	Novel Protein sim. GBank gi|82091|pir||A25494 - hydroxyproline-rich glycoprotein - tomato (fragment)		collagen	22278996, 29331822, 29331828, 264107, 264909, 264110, 265009, 264592, 264593, 60433356, 264288, 264693, 263974, 263976, 20281071, 60432113
491	80496412 (981, 982)	Novel Protein sim. GBank gi|2894206|emb|CAA17072| - (AL021840) hypothetical protein Rv3258c [Mycobacterium tuberculosis]		UNCLASSIFIED	264769
492	87421264 (983, 984)				264600
493	11692942 (985, 986)			UNCLASSIFIED	264638
494	87726604 (987, 988)	Novel Protein sim. GBank gi|5262605|emb|CAB45743.1| - (AL080150) hypothetical protein [Homo sapiens]		UNCLASSIFIED	264489, 35696286, 60432289, 29331828, 35696052, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265009, 264910, 33657402, 264762, 264764, 264768, 264769, 264688, 21906765, 21906769, 35695917, 265020, 264693, 33657109, 264629, 35696423, 35695855, 264634, 264638
495	80028599 (989, 990)	Novel Protein sim. GBank gi|2791517|emb|CAA16054| - (AL021246) hypothetical protein Rv2477c [Mycobacterium tuberculosis]	Contains protein domain (PF00005) - ABC transporter	transport	264602, 264682, 264638
496	78985624 (991, 992)	Novel Protein sim. GBank gi|230281|pdb|1R69| - 434 Repressor (Amino-Terminal Domain) (R1-69)	Contains protein domain (PF01381) - Helix-turn-helix		264601, 265021
497	78949661 (993, 994)	Novel Protein sim. GBank gi|129736|sp|P28225|PDXH_ECOLI - PYRIDOXAMINE 5'-PHOSPHATE OXIDASE (PNP/PMP OXIDASE)		oxidase	265006

498	88095488 (995, 996)	Novel Protein sim. GBank gi|1145789 (U41662) - neuroligin 2 [Rattus norvegicus]	Contains protein domain (PF00135) - Carboxylesterases	esterase	264259, 29331826, 35696052, 264508, 264509, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 265009, 264910, 264591, 33657402, 264758, 265010, 265011, 264600, 264601, 264605, 264683, 264764, 264766, 264767, 264768, 264687, 264769, 21906767, 33657023, 264693, 264628, 264629, 35696423, 264630, 264632, 264634, 264635, 264637, 264638, 264558, 264639, 18108385, 264563, 264564, 264565, 264566, 264567
499	20438222 (997, 998)	Novel Protein sim. GBank gi|97480|pir||S19739 - integral membrane protein - Rhodobacter capsulatus		UNCLASSIFIED	264605
500	11076810 (999, 1000)				264605
501	13418034 (1001, 1002)	Novel Protein sim. GBank gi|5708250|emb|CAB52363.1| - (AL109747) putative integral membrane protein [Streptomyces coelicolor A3(2)]		UNCLASSIFIED	264688
502	80021176 (1003, 1004)	Novel Protein sim. GBank gi|4468678|emb|CAB38132.1| - (AL035591) glucose-6-phosphate isomerase [Streptomyces coelicolor]	Contains protein domain (PF00342) - Phosphoglucose isomerase	isomerase	22278996, 265011, 264602, 264605, 264635
503	20264483 (1005, 1006)			UNCLASSIFIED	264564
504	10887321 (1007, 1008)			UNCLASSIFIED	264687
505	95003068 (1009, 1010)			UNCLASSIFIED	264369
506	16454292 (1011, 1012)	Novel Protein sim. GBank gi|4033509|sp|P02598|CALM_TETPY - CALMODULIN	Contains protein domain (PF00036) - EF hand	struct	265010
507	20451598 (1013, 1014)	Novel Protein sim. GBank gi|2501069|sp|Q46127|SYW_CLOLO - TRYPTOPHANYL-TRNA SYNTHETASE (TRYPTOPHAN--TRNA LIGASE) (TRPRS)		UNCLASSIFIED	264604
508	79841424 (1015, 1016)	Novel Protein sim. GBank gi|466068|sp|P34618|YO82_CAEEL - HYPOTHETICAL 33.8 KD PROTEIN ZK1236.2 IN CHROMOSOME III		UNCLASSIFIED	264908
509	11776386 (1017, 1018)				264638
510	83373465 (1019, 1020)			UNCLASSIFIED	264687, 264639
511	16525578 (1021, 1022)				265007
512	20399484 (1023, 1024)	Novel Protein sim. GBank gi|2497419|sp|P55635|Y4RB_RHISN - PUTATIVE INTEGRASE/RECOMBINASE Y4RB		UNCLASSIFIED	264565
513	79457404 (1025, 1026)	Novel Protein sim. GBank gi|1276897 (U41809) - cyclin J [Drosophila melanogaster]	Contains protein domain (PF00134) - Cyclin	cyclin	264683, 264689, 35696423, 264639
514	79813805 (1027, 1028)	Novel Protein sim. GBank gi|1184790 (U46068) - von Ebner minor salivary gland protein [Mus musculus]		UNCLASSIFIED	29331830, 264909
515	79462591 (1029, 1030)				22278999, 264690
516	9862020 (1031, 1032)	Novel Protein sim. GBank gi|2127400|pir||S65770 - maltooligosyltrehalose trehalohydrolase - Arthrobacter sp. (strain Q36)		amylase	264910

517	95292994 (1033, 1034)	Novel Protein sim. GBank gi|2983605 (AE000725) - ribose 5 phosphate isomerase B [Aquifex aeolicus]		isomerase	265018, 264605, 264764, 264766, 264687, 264691, 264565
518	8491831 (1035, 1036)	Novel Protein sim. GBank gi|854065|emb|CAA58337| - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264487
519	91677888 (1037, 1038)	Novel Protein sim. GBank gi|5689365|dbj|BAA83073.1| - (AB024075) B120 [Homo sapiens]	Contains protein domain (PF01388) - ARID DNA binding domain	dna_rna_bind	52644507, 22278997, 22278998, 60432049, 264259, 52645080, 29331824, 66714117, 60424269, 29331826, 35696052, 264905, 29331830, 66712502, 264511, 265007, 264591, 60432229, 33657402, 60433438, 21906754, 33109954, 52644296, 87168474, 87168559, 265017, 265018, 264604, 265019, 264681, 264448, 264369, 264288, 264685, 21906765, 21906766, 21906767, 21906769, 265021, 60170615, 33657023, 264692, 52645129, 33657109, 27486262, 27486264, 35695763, 18108370, 264629, 52644332, 56182323, 264639, 83373044, 18108385, 56526486, 60432113
520	79869188 (1039, 1040)				264769
521	11076821 (1041, 1042)	Novel Protein sim. GBank gi|1169126|sp|P46839|CTPA_MYCLE - CATION-TRANSPORTING P-TYPE ATPASE A		transport	264605
522	80435060 (1043, 1044)	Novel Protein sim. GBank gi|1172869|sp|P44331|RBSK_HAEIN - RIBOKINASE	Contains protein domain (PF00294) - pfkB family carbohydrate kinase	kinase	264905, 264768
523	18356013 (1045, 1046)	Novel Protein sim. GBank gi|2132243|pir||S61028 - hypothetical protein YPL236c - yeast (Saccharomyces cerevisiae)		UNCLASSIFIED	264629
524	80261805 (1047, 1048)	Novel Protein sim. GBank gi|4033608|dbj|BAA35136| - (AB012308) B2HC [Anthocidaris crassispina]		ATPase_associated	264092, 264596, 265011
525	79610046 (1049, 1050)				264907
526	36827630 (1051, 1052)	Novel Protein sim. GBank gi|4106610|emb|CAA21365| - (AL031866) ORF42, len=386 aa , similarity to an aminotransferase, in P95957 Sulfolobus solfataricus, (401 aa), 33.1% identity in 393 aa overlap, Fasta scores: opt:468, E(): 8.5e-24, in Q64602 R. norvegicus,(425 aa), 28.6% ident...		UNCLASSIFIED	264758
527	80504729 (1053, 1054)			UNCLASSIFIED	264769
528	65484134 (1055, 1056)			UNCLASSIFIED	56182575, 265017, 265018
529	17936810 (1057, 1058)	Novel Protein sim. GBank gi|731088|sp|P24215|UXUA_ECOLI - MANNONATE DEHYDRATASE (D-MANNONATE HYDROLASE)		hydrolase	265019
530	10887336 (1059, 1060)	Novel Protein sim. GBank gi|42144|emb|CAA25200| - (X00513) NusA protein (nusA) [Escherichia coli]		UNCLASSIFIED	264687
531	80226576 (1061, 1062)			UNCLASSIFIED	264555, 264556, 264557, 264558, 18108385
532	90933444 (1063, 1064)	Novel Protein sim. GBank gi|5262640|emb|CAB45758.1| - (AL080170) hypothetical protein [Homo sapiens]	Contains protein domain (PF00622) - SPRY domain	UNCLASSIFIED	264488, 264490, 264259, 264592, 264760, 265021, 264690, 263976, 264558

533	87761531 (1065, 1066)	Novel Protein sim. GBank gi|4883636|gb|AAD31593.1|AF11229 - (AF112299) integral inner nuclear membrane protein MAN1 [Homo sapiens]			264907, 264909, 264768, 35695917, 264630, 264555
534	82368264 (1067, 1068)	Novel Protein sim. GBank gi|2995352|emb|CAA04606.1| - (AJ001206) pep1 [Streptomyces coelicolor]		UNCLASSIFIED	264905, 265011, 264601, 264602, 264605, 264762, 264768, 265020, 264693, 264636
535	79641850 (1069, 1070)	Novel Protein sim. GBank gi|3878636|emb|CAA88953| - (Z49128) similar to cAMP-dependant protein kinase; cDNA EST EMBL:T00719 comes from this gene; cDNA EST yk465d8.3 comes from this gene; cDNA EST yk465d8.5 comes from this gene; cDNA EST yk492f4.3 comes from this gene; cDNA EST y...	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	ATPase_associated	264906
536	79907207 (1071, 1072)	Novel Protein sim. GBank gi|2495628|sp|P55757|YOHL_SERMA - HYPOTHETICAL 10.1 KD PROTEIN IN BIOA 5'REGION		reductase	18108376, 264905, 264906, 264907, 264909
537	94147448 (1073, 1074)				265008, 264605, 65274791
538	87821963 (1075, 1076)	Novel Protein sim. GBank gi|134920|sp|P21997|SSGP_VOLCA - SULFATED SURFACE GLYCOPROTEIN 185 (SSG 185)	Contains protein domain (PF00595) - PDZ domain (Also known as DHR or GLGF).	collagen	29331822, 29331824, 29331825, 29331826, 29331827, 264908, 52644045, 33657402, 265017, 264762, 264683, 264288, 264685, 21906765, 35695763, 264558, 60170394, 264559, 22279002
539	28396269 (1077, 1078)	Novel Protein sim. GBank gi|2498433|sp|Q12341|HAT1_YEAST - HISTONE ACETYLTRANSFERASE		histone	264602, 265019
540	79637077 (1079, 1080)				264693
541	87762268 (1081, 1082)	Novel Protein sim. GBank gi|3882241|dbj|BAA34480.1| - (AB018303) KIAA0760 protein [Homo sapiens]	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	18108394, 22278997, 22278998, 264259, 264112, 265009, 33657402, 55812038, 52646317, 265017, 21906765, 264693, 55811576, 264635, 56526486, 264566
542	95295836 (1083, 1084)	Novel Protein sim. GBank gi|5042272|emb|CAB44526.1| - (AL078618) nuoF, NADH dehydrogenase subunit [Streptomyces coelicolor]		dehydrogenase	264910, 265018, 264689, 264638, 264486
543	79796290 (1085, 1086)			UNCLASSIFIED	264602, 264908
544	20437191 (1087, 1088)	Novel Protein sim. GBank gi|2791398|emb|CAA15994| - (AL021184) hypothetical protein Rv1464 [Mycobacterium tuberculosis]		UNCLASSIFIED	264605
545	80434504 (1089, 1090)				264768, 264634, 264907, 264592, 264909
546	80249016 (1091, 1092)	Novel Protein sim. GBank gi|4887211|gb|AAD32237.1|AF14744 - (AF147449) penicillin binding protein 1B [Pseudomonas aeruginosa]			264600, 264602, 21906765
547	11077563 (1093, 1094)	Novel Protein sim. GBank gi|1350855|sp|P19176|RPOC_PSEPU - DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (TRANSCRIPTASE BETA' CHAIN) (RNA POLYMERASE BETA' SUBUNIT)		mapolymerase	264604
548	82114936 (1095, 1096)	Novel Protein sim. GBank gi|2330021 (AF019250) - kinesin-related protein; KRP; Costal2 [Drosophila melanogaster]		UNCLASSIFIED	264488, 264905, 264910, 264760, 264693, 264639, 264563, 264564

549	95421904 (1097, 1098)	Novel Protein sim. GBank gi|4337460|gb|AAD18133| - (AF056195) neuroblastoma-amplified protein [Homo sapiens]		UNCLASSIFIED	264488, 65274572, 18108398, 22278995, 22278996, 22278997, 22278998, 22278999, 264259, 29331824, 66714117, 29331825, 29331826, 35696052, 265007, 265008, 264910, 264592, 33657402, 33109954, 265017, 265018, 265019, 18108351, 264448, 264764, 264369, 264288, 264766, 264686, 264688, 21906765, 21906766, 21906767, 21906768, 21906769, 265020, 264691, 33657023, 264692, 264693, 65274620, 52645129, 33657109, 27486261, 27486262, 27486264, 33657349, 55811576, 18108387, 60432113, 22279002
550	10886616 (1099, 1100)				264688
551	80439990 (1101, 1102)	Novel Protein sim. GBank gi|3122893|sp|P94985|SYFB_MYCTU - PHENYLALANYL-TRNA SYNTHETASE BETA CHAIN (PHENYLALANINE--TRNA LIGASE BETA CHAIN) (PHERS)		UNCLASSIFIED	264908, 264909, 264768
552	94672870 (1103, 1104)			UNCLASSIFIED	264689, 264639, 264563
553	80106002 (1105, 1106)	Novel Protein sim. GBank gi|552087 (M33753) - crumbs protein [Drosophila melanogaster]	Contains protein domain (PF00008) - EGF-like domain	glycoprotein	55811957, 264628
554	79618379 (1107, 1108)	Novel Protein sim. GBank gi|5019771|gb|AAD37857.1|AF13326 - (AF133263) histidine protein kinase-response regulator hybrid protein CvgSY [Pseudomonas syringae pv. syringae]		kinase	264906
555	78996347 (1109, 1110)	Novel Protein sim. GBank gi|131515|sp|P02908|PTGA_SALTY - PTS SYSTEM, GLUCOSE-SPECIFIC IIA COMPONENT (EIIA-GLC) (GLUCOSE-PERMEASE IIA COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, A COMPONENT) (EIII-GLC)	Contains protein domain (PF00358) - phosphoenolpyruvate-dependent sugar phosphotransferase system, EIIA 1	transport	264762
556	20457127 (1111, 1112)	Novel Protein sim. GBank gi|3914014|sp|P96380|MFD_MYCTU - TRANSCRIPTION-REPAIR COUPLING FACTOR (TRCF)		transcriptfactor	264508, 264605, 264559
557	19523405 (1113, 1114)	Novel Protein sim. GBank gi|5042273|emb|CAB44527.1| - (AL078618) nuoE, NADH dehydrogenase subunit [Streptomyces coelicolor]		dehydrogenase	264488
558	20724429 (1115, 1116)	Novel Protein sim. GBank gi|1170933|sp|P45331|METE_HAEIN - 5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE--HOMOCYSTEINE METHYLTRANSFERASE (METHIONINE SYNTHASE, VITAMIN-B12 INDEPENDENT ISOZYME) (COBALAMIN-INDEPENDENT METHIONINE SYNTHASE)		UNCLASSIFIED	264602
559	80084353 (1117, 1118)	Novel Protein sim. GBank gi|4980567|gb|AAD35173.1|AE00169 - (AE001694) iron(III) ABC transporter, permease protein [Thermotoga maritima]		UNCLASSIFIED	264634

560	80066533 (1119, 1120)	Novel Protein sim. GBank gi|2492595|sp|Q53193|Y4TR_RHISN - PROBABLE PEPTIDE ABC TRANSPORTER ATP-BINDING PROTEIN Y4TR	Contains protein domain (PF00005) - ABC transporter	transport	18108396, 264906, 264602, 264604, 18108374
561	20293187 (1121, 1122)			UNCLASSIFIED	264600
562	11698161 (1123, 1124)			UNCLASSIFIED	264689
563	79761420 (1125, 1126)	Novel Protein sim. GBank gi|4104925 (AF042276) - poly(hydroxyalcanoate) granule associated protein GA2 [Pseudomonas putida]		UNCLASSIFIED	264910, 264691
564	56716390 (1127, 1128)	Novel Protein sim. GBank gi|2792310 (AF040570) - unknown [Amycolatopsis mediterranei]		dehydrogenase	264592
565	56465618 (1129, 1130)	Novel Protein sim. GBank gi|3449294|db|BAA32462| - (AB011532) MEGF6 [Rattus norvegicus]	Contains protein domain (PF00008) - EGF-like domain	synthase	265010
566	94323888 (1131, 1132)	Novel Protein sim. GBank gi|4539568|emb|CAB38487.1| - (AL035636) putative helicase [Streptomyces coelicolor]		helicase	264909, 264510, 265008, 264910, 264758, 264600, 264602, 264604, 264605, 264768, 264687, 264689, 35695917, 264693, 65274620, 264488
567	79560955 (1133, 1134)			UNCLASSIFIED	264681, 264691, 264593
568	94681793 (1135, 1136)	Novel Protein sim. GBank gi|100506|pir||S17455 - Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP+) (EC 1.1.1.40) - Flaveria trinervia (fragment)	Contains protein domain (PF00390) - Malic enzyme	dehydrogenase	264689
569	39506897 (1137, 1138)	Novel Protein sim. GBank gi|3915843|sp|O31212|RS2_STRCO - 30S RIBOSOMAL PROTEIN S2	Contains protein domain (PF00318) - Ribosomal protein S2	ribosomalprot	264565
570	78375927 (1139, 1140)			UNCLASSIFIED	18108376, 18108387, 264565
571	79793961 (1141, 1142)	Novel Protein sim. GBank gi|115122|sp|P21627|BRAD_PSEAE - HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PROTEIN BRAD		transport	264907, 264909
572	36996838 (1143, 1144)			UNCLASSIFIED	264762
573	20715521 (1145, 1146)	Novel Protein sim. GBank gi|4539223|emb|CAB39881.1| - (AL049497) putative integral membrane protein [Streptomyces coelicolor]		UNCLASSIFIED	265007, 264601
574	13521592 (1147, 1148)				264636
575	13076416 (1149, 1150)	Novel Protein sim. GBank gi|118794|sp|P10443|DP3A_ECOLI - DNA POLYMERASE III, ALPHA CHAIN		polymerase	264687
576	20482246 (1151, 1152)	Novel Protein sim. GBank gi|5457625|emb|CAB49116.1| - (AJ248283) PAB2227 [Pyrococcus abyssi]			264605
577	66727102 (1153, 1154)	Novel Protein sim. GBank gi|5042274|emb|CAB44528.1| - (AL078618) nuoD, NADH dehydrogenase subunit [Streptomyces coelicolor]	Contains protein domain (PF00346) - Respiratory-chain NADH dehydrogenase, 49 Kd subunit	dehydrogenase	35696052, 264636
578	11804477 (1155, 1156)				264638
579	11794723 (1157, 1158)	Novel Protein sim. GBank gi|1723081|sp|Q11046|Y089_MYCTU - HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN CY50.09		transport	264682, 264556

580	80059417 (1159, 1160)				22278999, 35696052, 264555, 264556, 264558
581	79230833 (1161, 1162)			UNCLASSIFIED	265008, 264564
582	80049617 (1163, 1164)	Novel Protein sim. GBank gi|3243131 (AF045777) - titin [Drosophila melanogaster]	Contains protein domain (PF00047) - Immunoglobulin domain	struct	265021, 264555, 264557
583	79321392 (1165, 1166)	Novel Protein sim. GBank gi|2501162|sp|P77726|YAJR_ECOLI - HYPOTHETICAL 49.0 KD PROTEIN IN ABPA-CYOE INTERGENIC REGION		transport	264594
584	79845024 (1167, 1168)			UNCLASSIFIED	264488, 264906, 264766, 264687, 35696423
585	79581454 (1169, 1170)	Novel Protein sim. GBank gi|3882221|dbj|BAA34470.1| - (AB018293) KIAA0750 protein [Homo sapiens]		UNCLASSIFIED	265018, 264684, 21906769
586	38277486 (1171, 1172)			UNCLASSIFIED	264908, 265007
587	80497359 (1173, 1174)	Novel Protein sim. GBank gi|4467250|emb|CAB37575| - (AL035569) probable Glu-tRNA Gln amidotransferase subunit [Streptomyces coelicolor]		hydrolase	264600, 264602, 264605, 264769, 264690, 264557
588	79557239 (1175, 1176)	Novel Protein sim. GBank gi|5689519|dbj|BAA83043.1| - (AB029014) KIAA1091 protein [Homo sapiens]		UNCLASSIFIED	265020, 264692
589	79805828 (1177, 1178)			UNCLASSIFIED	22278996, 264907, 264909, 264510, 265009, 265010, 264687, 264769, 35695917, 18108376, 264634, 264636, 264638
590	79815629 (1179, 1180)			UNCLASSIFIED	264906, 264909
591	10313540 (1181, 1182)	Novel Protein sim. GBank gi|2143293|emb|CAB09390| - (Z95972) rpoB [Mycobacterium tuberculosis]		mapolymerase	264691
592	13889767 (1183, 1184)			MHC	263972
593	82348699 (1185, 1186)	Novel Protein sim. GBank gi|4511983|gb|AAD21543.1| - (AF088896) electrotransfer ubiquinone oxidoreductase [Zymomonas mobilis]		dehydrogenase	264511, 264762, 264769, 264486
594	20212392 (1187, 1188)	Novel Protein sim. GBank gi|1272368 (U51896) - LfgE [Vibrio parahaemolyticus]		UNCLASSIFIED	264605
595	10064064 (1189, 1190)	Novel Protein sim. GBank gi|131490|sp|P20966|PTFB_ECOLI - PTS SYSTEM, FRUCTOSE-SPECIFIC IIBC COMPONENT (EIIBC-FRU) (FRUCTOSE-PERMEASE IIBC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, BC COMPONENT) (EII-FRU)			264769
596	13085170 (1191, 1192)			UNCLASSIFIED	264636
597	80259003 (1193, 1194)			UNCLASSIFIED	264592
598	94140216 (1195, 1196)			UNCLASSIFIED	264758, 55810764, 264555, 264556, 264637, 83373044
599	20385137 (1197, 1198)	Novel Protein sim. GBank gi|125329|sp|P04951|KDSB_ECOLI - 3-DEOXY-MANNO-OCTULOSONATE CYTIDYLYLTRANSFERASE (CMP-KDO SYNTHETASE) (CMP-2-KETO-3-DEOXYOCTULOSONIC ACID SYNTHETASE) (CKS)		UNCLASSIFIED	264603
600	10357663 (1199, 1200)				264906
601	79610404 (1201, 1202)	Novel Protein sim. GBank gi|2127414|pir||S60064 - hypothetical protein 2 - Corynebacterium glutamicum		UNCLASSIFIED	264510

602	79250602 (1203, 1204)	Novel Protein sim. GBank gi|3522961|gb|AAC34243.1| - (AC004411) putative pto kinase [Arabidopsis thaliana]	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	265007
603	11466067 (1205, 1206)			UNCLASSIFIED	264595
604	81675420 (1207, 1208)				264758
605	20436657 (1209, 1210)	Novel Protein sim. GBank gi|1175322|sp|P44917|Y883_HAEIN - HYPOTHETICAL PROTEIN HI0883		UNCLASSIFIED	264605
606	80334582 (1211, 1212)	Novel Protein sim. GBank gi|5020264|gb|AAD38043.1|AF15136 - (AF151383) Cdc42 GTPase-activating protein [Mus musculus]		UNCLASSIFIED	264764
607	95361506 (1213, 1214)	Novel Protein sim. GBank gi|188864 (M74027) - mucin [Homo sapiens]		UNCLASSIFIED	264508, 264906, 85658542, 264682, 264687, 264689, 264534, 18108376, 35696423, 264636, 264555, 264638
608	11810888 (1215, 1216)			UNCLASSIFIED	264682
609	80064775 (1217, 1218)	Novel Protein sim. GBank gi|2496701|sp|P55552|Y4LL_RHISN - HYPOTHETICAL 91.8 KD PROTEIN Y4LL	Contains protein domain (PF00989) - PAS domain	UNCLASSIFIED	264605
610	79629413 (1219, 1220)				264692
611	87586205 (1221, 1222)				264508, 264905, 264907, 264908, 264909, 264511, 264910, 264758, 264604, 264684, 264766, 264689, 264692, 264628, 264635, 264636, 264637, 264558
612	95287851 (1223, 1224)	Novel Protein sim. GBank gi|1877366|emb|CAB07118| - (Z92772) recD [Mycobacterium tuberculosis]	Contains protein domain (PF01443) - Viral (Superfamily 1) RNA helicase	nuclease	264600, 264601, 264604, 264769, 264558, 264565
613	7523475 (1225, 1226)			UNCLASSIFIED	264369
614	79969348 (1227, 1228)	Novel Protein sim. GBank gi|5114231|gb|AAD40238.1|AF13670 - (AF136709) histidine kinase YycG [Staphylococcus aureus]		kinase	18108372, 264563
615	39586996 (1229, 1230)	Novel Protein sim. GBank gi|1339950|dbj|BAA12741| - (D85230) large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]		synthase	264600, 264602, 264629
616	20465331 (1231, 1232)	Novel Protein sim. GBank gi|544367|sp|P35673|GALE_ERWAM - UDP-GLUCOSE 4-EPIMERASE (GALACTOWALDENASE) (UDP-GALACTOSE 4-EPIMERASE)		isomerase	264605
617	91227222 (1233, 1234)	Novel Protein sim. GBank gi|2498097|sp|Q60769|TNP3_MOUSE - TUMOR NECROSIS FACTOR, ALPHA-INDUCED PROTEIN 3 (PUTATIVE DNA BINDING PROTEIN A20) (ZINC FINGER PROTEIN A20)	Contains protein domain (PF00641) - Zn-finger in Ran binding protein and others.	tnf	52645156, 21906765, 35696423, 21906768, 21906769, 22278994, 35696286, 22278996, 265020, 265021, 265007, 265008, 264636, 52644150, 33657023, 264692, 264693, 29331822, 29331824, 55812038, 83373044, 56182181, 60424269, 66714117, 29331825, 33657109, 29331826, 33657182, 29331827, 35696052, 29331828, 27486262, 33657349, 56526486, 265018, 265019, 22279002, 264482, 264448, 29331830, 66712502, 264909

618	20632843 (1235, 1236)	Novel Protein sim. GBank gi|5459388|emb|CAB50746.1| - (AL096839) putative aminotransferase [Streptomyces coelicolor]		isomerase	264603
619	91227224 (1237, 1238)				56994075, 29331826, 33656970, 265008, 33657402, 33109954, 87168559, 264448, 18108374, 83373044
620	81183143 (1239, 1240)	Novel Protein sim. GBank gi|464335|sp|Q05922|DUS2_MOUSE - DUAL SPECIFICITY PROTEIN PHOSPHATASE 2 (DUAL SPECIFICITY PROTEIN PHOSPHATASE PAC-1)		phosphatase	29146498, 264758, 264369, 29148627
621	80239251 (1241, 1242)			UNCLASSIFIED	264556, 264558, 264639
622	20456427 (1243, 1244)	Novel Protein sim. GBank gi|2633557|emb|CAB13060| - (Z99110) yjdF [Bacillus subtilis]		UNCLASSIFIED	264605
623	10131798 (1245, 1246)	Novel Protein sim. GBank gi|1857710|gb|AAB48482| - (U87224) contactin associated protein [Rattus norvegicus]	Contains protein domain (PF00054) - Laminin G domain	laminin	264906
624	19534127 (1247, 1248)	Novel Protein sim. GBank gi|1705703|sp|P52225|CCMF_PSEFL - CYTOCHROME C-TYPE BIOGENESIS PROTEIN CYCK		cytochrome	264596
625	13084619 (1249, 1250)	Novel Protein sim. GBank gi|2894252|emb|CAA17114.1| - (AL021841) hypothetical protein Rv3342 [Mycobacterium tuberculosis]		UNCLASSIFIED	264688
626	88062603 (1251, 1252)	Novel Protein sim. GBank gi|416592|sp|P32323|AGA1_YEAST - A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR		UNCLASSIFIED	29331822, 264905, 264908, 33657023, 33657109, 264558
627	80255457 (1253, 1254)	Novel Protein sim. GBank gi|3098418 (AF040944) - P140 [Mus musculus]		UNCLASSIFIED	18108394, 264112, 264593, 265022, 264635
628	80077096 (1255, 1256)	Novel Protein sim. GBank gi|1711543|sp|P50526|SSP1_SCHPO - SERINE/THREONINE-PROTEIN KINASE SSP1	Contains protein domain (PF00069) - Eukaryotic protein kinase domain	kinase	264600
629	79851602 (1257, 1258)	Novel Protein sim. GBank gi|1143204 (U34305) - ORF2; Method: conceptual translation supplied by author. [Shigella sonnei]		isomerase	264906, 264907
630	39565156 (1259, 1260)	Novel Protein sim. GBank gi|3236368 (AF064748) - S3-12 [Mus musculus]		UNCLASSIFIED	264490
631	20598718 (1261, 1262)	Novel Protein sim. GBank gi|140687|sp|P11666|YGGB_ECOLI - HYPOTHETICAL 30.9 KD PROTEIN IN SBM-FBA INTERGENIC REGION (ORF 4) (F286)			263978
632	27843890 (1263, 1264)			UNCLASSIFIED	264906, 264600, 264605, 264769, 264689, 264486
633	80477772 (1265, 1266)			UNCLASSIFIED	264769
634	17938806 (1267, 1268)				265019
635	79574506 (1269, 1270)			UNCLASSIFIED	264689
636	79910981 (1271, 1272)			UNCLASSIFIED	264596, 264762, 264693

637	82455796 (1273, 1274)	Novel Protein sim. GBank gi|2326739|emb|CAB10953| - (Z98268) recN [Mycobacterium tuberculosis]		nuclease	264906, 264907, 264510, 264511, 264601, 264602, 264603, 264604, 264605, 18108351, 264762, 264766, 264687, 264769, 264689, 35695917, 264693, 264634, 264638, 264639, 264559, 18108385
638	14997457 (1275, 1276)	Novel Protein sim. GBank gi|4678662|emb|CAB41074.1| - (AL049645) putative large ATP-binding protein [Streptomyces coelicolor]			264636
639	80204210 (1277, 1278)	Novel Protein sim. GBank gi|4589628|dbj|BAA76836.1| - (AB023209) KIAA0992 protein [Homo sapiens]		struct	264112, 263974
640	17929579 (1279, 1280)	Novel Protein sim. GBank gi|1432083 (U60981) - homolog to Skp1p, an evolutionarily conserved kinetochore protein in budding yeast [Arabidopsis thaliana]	Contains protein domain (PF01466) - Skp1 family	rnapolymerase	265009, 265010
641	79636398 (1281, 1282)			UNCLASSIFIED	264693
642	19898737 (1283, 1284)			UNCLASSIFIED	264565
643	81516220 (1285, 1286)			UNCLASSIFIED	264906, 264908, 264758, 264288, 264632, 264635, 264639, 264564
644	11751367 (1287, 1288)			UNCLASSIFIED	264684
645	95010907 (1289, 1290)			UNCLASSIFIED	264906, 264762, 264693, 264639, 264559
646	80069083 (1291, 1292)				264595, 264566
647	80257085 (1293, 1294)	Novel Protein sim. GBank gi|4507613|ref|NP_003738.1|pTNKS - TANKYRASE	Contains protein domain (PF00023) - Ank repeat	transcriptfactor	264909, 264591
648	80077428 (1295, 1296)	Novel Protein sim. GBank gi|1044963|bbs|169646 - protamine [Monodonta turbinata, gonads, Peptide, 106 aa]		UNCLASSIFIED	264600
649	80247447 (1297, 1298)			UNCLASSIFIED	263978
650	11798316 (1299, 1300)			UNCLASSIFIED	264686
651	11776932 (1301, 1302)	Novel Protein sim. GBank gi|1346916|sp|P12283|PURA_ECOLI - ADENYLOSUCCINATE SYNTHETASE (IMP--ASPARTATE LIGASE)			264602, 264638
652	85516704 (1303, 1304)			UNCLASSIFIED	264905, 264907, 264909, 263978, 264637
653	82124947 (1305, 1306)	Novel Protein sim. GBank gi|1722977|sp|Q10638|Y03C_MYCTU - HYPOTHETICAL 82.8 KD PROTEIN CY130.12C		UNCLASSIFIED	22278996, 264510, 264511, 264512, 264593, 21906754, 264603, 264760, 18108376, 264556
654	95010589 (1307, 1308)			UNCLASSIFIED	264906, 264595, 264632
655	79320692 (1309, 1310)	Novel Protein sim. GBank gi|130327|sp|P26647|PLSC_ECOLI - 1-ACYL-SN-GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE (1-AGP ACYLTRANSFERASE) (1-AGPAT) (LYSOPHOSPHATIDIC ACID ACYLTRANSFERASE) (LPAAT)	Contains protein domain (PF01553) - Acyltransferase	transferase	264592
656	80416739 (1311, 1312)			UNCLASSIFIED	264602, 264605, 264766, 264691
657	20811010 (1313, 1314)			UNCLASSIFIED	264557, 264558

658	87761915 (1315, 1316)	Novel Protein sim. GBank gi|5689493|dbj|BAA83030.1| - (AB029001) KIAA1078 protein [Homo sapiens]		UNCLASSIFIED	22278996, 60432049, 29331822, 29331824, 29331828, 265007, 265009, 33657402, 33657084, 265017, 264448, 21906765, 21906766, 263967, 20281149, 18108370, 18108374, 264482
659	87718663 (1317, 1318)	Novel Protein sim. GBank gi|2137872|pir||I48724 - zinc finger protein PZF - mouse	Contains protein domain (PF00096) - Zinc finger, C2H2 type	transcriptfactor	22278999, 60432049, 66714117, 29331827, 265007, 264766, 56181562, 18108359, 18108365, 18108370, 18108381
660	81897922 (1319, 1320)			UNCLASSIFIED	264757
661	80026023 (1321, 1322)	Novel Protein sim. GBank gi|134180|sp|P15401|SACY_BACSU - LEVANSUCRASE AND SUCRASE SYNTHESIS OPERON ANTITERMINATOR	Contains protein domain (PF00874) - Transcriptional antiterminator bglG family	UNCLASSIFIED	264510, 265009, 264600, 264602, 264603, 264604, 264605, 32833986, 18108376, 264636, 18108387, 22279000
662	20463731 (1323, 1324)	Novel Protein sim. GBank gi|4545229|gb|AAD22450.1|AF11618 - (AF116183) SecA homolog [Actinobacillus actinomycetemcomitans]		UNCLASSIFIED	264605
663	20628080 (1325, 1326)	Novel Protein sim. GBank gi|5689250|dbj|BAA82881.1| - (AB024335) similar to orf5 [Comamonas testosteroni]		dehydrogenase	264605
664	80508512 (1327, 1328)	Novel Protein sim. GBank gi|1652848|dbj|BAA17766| - (D90909) DNA photolyase [Synechocystis sp.]		UNCLASSIFIED	264769
665	80079053 (1329, 1330)	Novel Protein sim. GBank gi|116841|sp|P21640|COBJ_PSEDE - PRECORRIN-3B C17 METHYLTRANSFERASE (PRECORRIN-3 METHYLTRANSFERASE) (PRECORRIN-3 METHYLASE)		isomerase	264600
666	79603142 (1331, 1332)	Novel Protein sim. GBank gi|3261829|emb|CAB10927| - (Z98260) hypothetical protein Rv1230c [Mycobacterium tuberculosis]		glycoprotein	264907, 265007
667	94631802 (1333, 1334)	Novel Protein sim. GBank gi|5688851|dbj|BAA82702.1| - (AB017438) Orf5 [Streptomyces coelicolor]		UNCLASSIFIED	264689, 264602, 264593
668	82051891 (1335, 1336)	Novel Protein sim. GBank gi|3581853|emb|CAA20809| - (AL031541) 50S ribosomal protein L20 [Streptomyces coelicolor]	Contains protein domain (PF00453) - Ribosomal protein L20	ribosomalprot	264905, 264906, 264908, 264600, 264601, 264603, 264605, 264760, 264689, 264636, 264638, 264639
669	12967154 (1337, 1338)			UNCLASSIFIED	264637
670	80238549 (1339, 1340)	Novel Protein sim. GBank gi|2582531 (AF026444) - 2-isopropylmalate synthase [Streptomyces coelicolor]		synthase	264905, 264906, 264908, 264601, 264762, 264766, 264689, 264638, 18108385, 264486
671	79601368 (1341, 1342)		Contains protein domain (PF00023) - Ank repeat	UNCLASSIFIED	264690, 264692, 264693, 264636, 18108387
672	79834371 (1343, 1344)	Novel Protein sim. GBank gi|2114430 (U92703) - Olf-1/EBF-like-3 transcription factor [Mus musculus]		transcriptfactor	264910, 265017
673	82285798 (1345, 1346)	Novel Protein sim. GBank gi|4589285|gb|AAD26430.1|AF13515 - (AF135154) ferric alcaligin siderophore receptor [Bordetella pertussis]			264759
674	79199259 (1347, 1348)			UNCLASSIFIED	264629

675	87895870 (1349, 1350)	Novel Protein sim. GBank gi|4980755|gb|AAD35347.1|AE00170 - (AE001708) D-alanine--D-alanine ligase [Thermotoga maritima]	Contains protein domain (PF01820) - D-ala D-ala ligase	UNCLASSIFIED	264488, 22278999, 66714117, 264508, 264511, 265008, 60433438, 264600, 264601, 264602, 264603, 264604, 264605, 264762, 264687, 264769, 60431602, 18108374, 264636, 264638
676	78899607 (1351, 1352)	Novel Protein sim. GBank gi|1723566|sp|Q10479|YDF7_SCHPO - PUTATIVE GLUCOSYLTRANSFERASE C17C9.07			265010
677	21644312 (1353, 1354)	Novel Protein sim. GBank gi|687208 (U03976) - dynein heavy chain isotype 5C [Tripneustes gratilla]		ATPase_associated	264591, 264632
678	84225200 (1355, 1356)	Novel Protein sim. GBank gi|1586274|prf||2203365A - laminin alpha5 [Mus musculus]	Contains protein domain (PF00053) - Laminin EGF-like (Domains III and V)	laminin	264758, 264682, 264557
679	79868855 (1357, 1358)	Novel Protein sim. GBank gi|3928723|emb|CAA22219| - (AL034355) putative ABC transporter [Streptomyces coelicolor]		UNCLASSIFIED	22278998, 264693
680	20726424 (1359, 1360)				264600, 264602
681	94322017 (1361, 1362)	Novel Protein sim. GBank gi|5174493|ref|NP_006050.1|pLAMC - laminin, gamma 3	Contains protein domain (PF00053) - Laminin EGF-like (Domains III and V)	laminin	264102, 264907, 264908, 265006, 264693, 263972, 83373044, 264566
682	11392476 (1363, 1364)			UNCLASSIFIED	264595
683	80083680 (1365, 1366)	Novel Protein sim. GBank gi|4758208|ref|NP_004081.1|pDUSP - dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	Contains protein domain (PF00782) - Dual specificity phosphatase, catalytic domain	phosphatase	264634
684	20465367 (1367, 1368)	Novel Protein sim. GBank gi|5420387|emb|CAB46679.1| - (AJ243459) proteophosphoglycan [Leishmania major]			264605
685	80246735 (1369, 1370)				264909, 263967, 263981
686	79208608 (1371, 1372)				264631
687	80085629 (1373, 1374)				264693, 264635
688	79853412 (1375, 1376)	Novel Protein sim. GBank gi|2688962 (AF027768) - LspA [Serratia marcescens]		peptidase	264907, 264638
689	88064256 (1377, 1378)	Novel Protein sim. GBank gi|3046931 (AF049330) - PPAR gamma coactivator [Mus musculus]	Contains protein domain (PF00076) - RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	UNCLASSIFIED	264906, 264907, 265007, 265009, 60433438, 21906754, 264760, 18108358, 21906766, 21906769, 265021, 18108361, 263974, 18108379, 264557, 18108385, 22279002
690	80389750 (1379, 1380)	Novel Protein sim. GBank gi|2498941|sp|Q15428|SP62_HUMAN - SPLICEOSOME ASSOCIATED PROTEIN 62 (SAP 62) (SF3A66)		UNCLASSIFIED	264510, 264511, 264764, 264769
691	81854392 (1381, 1382)			UNCLASSIFIED	264757
692	83608936 (1383, 1384)	Novel Protein sim. GBank gi|5420387|emb|CAB46679.1| - (AJ243459) proteophosphoglycan [Leishmania major]	Contains protein domain (PF00097) - Zinc finger, C3HC4 type (RING finger)	UNCLASSIFIED	55812038, 55811957, 265018, 55811150, 18108351, 264908, 60431528, 264594
693	79586116 (1385, 1386)	Novel Protein sim. GBank gi|854065|emb|CAA58337| - (X83413) U88 [Human herpesvirus 6]		UNCLASSIFIED	264635
694	82455983 (1387, 1388)	Novel Protein sim. GBank gi|267327|sp|Q01033|VG48_HSVSA - HYPOTHETICAL GENE 48 PROTEIN			22278996, 264510, 264602, 264603, 264762, 264687, 264769, 264693

695	94147849 (1389, 1390)	Novel Protein sim. GBank gi|4468339|emb|CAB38059.1| - (AJ010901) MUC4 [Homo sapiens]	Contains protein domain (PF00094) - von Willebrand factor type D domain	UNCLASSIFIED	56182575, 264509, 264905, 264907, 29331830, 264908, 264909, 264511, 265007, 264910, 264758, 264764, 264288, 65274791
696	79830982 (1391, 1392)	Novel Protein sim. GBank gi|2649950 (AE001058) - glutamine ABC transporter, ATP-binding protein (glnQ) [Archaeoglobus fulgidus]	Contains protein domain (PF00005) - ABC transporter	transport	264905, 264595
697	11767889 (1393, 1394)	Novel Protein sim. GBank gi|1731343|sp|Q10694|YY25_MYCTU - HYPOTHETICAL 24.4 KD PROTEIN CY49.25	Contains protein domain (PF01836) - Transposase	UNCLASSIFIED	264682
698	66695862 (1395, 1396)			UNCLASSIFIED	264688, 35695917
699	79582558 (1397, 1398)			UNCLASSIFIED	264682
700	79639098 (1399, 1400)				264693
701	80230242 (1401, 1402)	Novel Protein sim. GBank gi|1001236|dbj|BAA10477| - (D64003) hypothetical protein [Synechocystis sp.]		UNCLASSIFIED	264488, 264510, 264511, 264602, 264605, 264689
702	79814789 (1403, 1404)			UNCLASSIFIED	264909
703	20446820 (1405, 1406)	Novel Protein sim. GBank gi|2498935|sp|Q46338|SOXG_CORSP - SARCOSINE OXIDASE GAMMA SUBUNIT		oxidase	264604
704	94312224 (1407, 1408)	Novel Protein sim. GBank gi|3150513 (AF067219) - contains similarity to the kelch/MIPP family [Caenorhabditis elegans]	Contains protein domain (PF01344) - Kelch motif	UNCLASSIFIED	264288, 56181562, 33657109, 264629, 55811576
705	17932141 (1409, 1410)	Novel Protein sim. GBank gi|421091|pir||S30730 - hypothetical protein o206 - Escherichia coli		UNCLASSIFIED	265006
706	20288062 (1411, 1412)	Novel Protein sim. GBank gi|3024872|sp|Q55790|Y074_SYNY3 - HYPOTHETICAL 52.8 KD PROTEIN SLR0074			264600
707	20638065 (1413, 1414)	Novel Protein sim. GBank gi|3420608|gb|AAC31907.1| - (AF075709) ABC transporter ATP-binding subunit [Pseudomonas putida]		transport	264603
708	20708292 (1415, 1416)				264601, 264692
709	88001439 (1417, 1418)	Novel Protein sim. GBank gi|3649741|emb|CAA03985| - (AJ000281) mucin [Homo sapiens]		struct	18108398, 264637, 264908, 264909
710	11356683 (1419, 1420)	Novel Protein sim. GBank gi|3080425|emb|CAA18744.1| - (AL022604) putative protein [Arabidopsis thaliana]			264369
711	17931418 (1421, 1422)				265019
712	80258164 (1423, 1424)	Novel Protein sim. GBank gi|4758686|ref|NP_002323.1|pLRP1 - low density lipoprotein related protein 1 (alpha-2-macroglobulin receptor)	Contains protein domain (PF00058) - Low-density lipoprotein receptor repeat class B	apolipoprotein	264591
713	79263126 (1425, 1426)	Novel Protein sim. GBank gi|1703266|sp|Q11056|AMI2_MYCTU - PUTATIVE AMIDASE CY50.19C		hydrolase	264906, 264907
714	27847651 (1427, 1428)	Novel Protein sim. GBank gi|4502351|ref|NP_001692.1|pBAAT - bile acid Coenzyme A: amino acid N-acyltransferase; glycine N-choloyltransferase			264508, 264555

WO 00/58473

PCT/US00/08621

715	79639423 (1429, 1430)	Novel Protein sim. GBank gi|1789035 (AE000352) - orf, hypothetical protein [Escherichia coli]		UNCLASSIFIED	264907
716	79559072 (1431, 1432)				264692
717	79491842 (1433, 1434)	Novel Protein sim. GBank gi|2494074|sp|P55653|GABD_RHISN - PROBABLE SUCCINATE-SEMIALDEHYDE DEHYDROGENASE (NADP+) (SSDH)		dehydrogenase	264636
718	94319658 (1435, 1436)	Novel Protein sim. GBank gi|3873679|emb|CAA94886| - (Z71178) similar to pro-collagen domains; cDNA EST EMBL:D27978 comes from this gene; cDNA EST EMBL:D27977 comes from this gene; cDNA EST EMBL:D34199 comes from this gene; cDNA EST EMBL:D64392 comes from this gene; cDNA EST EMBL...	Contains protein domain (PF00093) - von Willebrand factor type C domain	kinase	18108392, 22278994, 22278998, 265008, 265018, 264681, 18108354, 264684, 264685, 264686, 264687, 264689, 21906769, 18108361, 264691, 264692, 55810764, 264635, 18108381, 18108382, 83373044, 18108388
719	17679564 (1437, 1438)	Novel Protein sim. GBank gi|2104302|emb|CAB08631| - (Z95387) hypothetical protein Rv2611c [Mycobacterium tuberculosis]		UNCLASSIFIED	265011
720	79841684 (1439, 1440)				264908
721	15020180 (1441, 1442)	Novel Protein sim. GBank gi|123530|sp|P04929|HRPX_PLALO - HISTIDINE-RICH GLYCOPROTEIN PRECURSOR		UNCLASSIFIED	264629
722	9862603 (1443, 1444)	Novel Protein sim. GBank gi|498253 (U02372) - integrase [Vibrio cholerae]			264910
723	19755599 (1445, 1446)	Novel Protein sim. GBank gi|2253054|emb|CAB10705| - (Z97559) hypothetical protein Rv2114 [Mycobacterium tuberculosis]		UNCLASSIFIED	264691
724	10126494 (1447, 1448)	Novel Protein sim. GBank gi|4063015 (AF083061) - protease PrtA [Pseudomonas fluorescens]	Contains protein domain (PF00353) - Hemolysin-type calcium-binding proteins	protease	264909
725	79878679 (1449, 1450)			UNCLASSIFIED	264905, 264907
726	13086282 (1451, 1452)			UNCLASSIFIED	264636
727	13522872 (1453, 1454)				264634
728	20268471 (1455, 1456)	Novel Protein sim. GBank gi|2633910|emb|CAB13411| - (Z99112) similar to hypothetical proteins [Bacillus subtilis]			264567
729	11293753 (1457, 1458)			UNCLASSIFIED	264490
730	19900373 (1459, 1460)	Novel Protein sim. GBank gi|2494660|sp|Q45291|GALE_BRELA - UDP-GLUCOSE 4-EPIMERASE (GALACTOWALDENASE) (UDP-GALACTOSE 4-EPIMERASE)		isomerase	264564
731	80058750 (1461, 1462)	Novel Protein sim. GBank gi|1146192 (L47838) - putative [Bacillus subtilis]		UNCLASSIFIED	264605
732	80258175 (1463, 1464)	Novel Protein sim. GBank gi|1168396|sp|P46681|AIP2_YEAST - ACTIN INTERACTING PROTEIN 2		struct	264591, 264594, 264595
733	20446839 (1465, 1466)			UNCLASSIFIED	264604
734	20435987 (1467, 1468)	Novel Protein sim. GBank gi|3184080|emb|CAA19336| - (AL023781) hypothetical protein [Schizosaccharomyces pombe]		ubiquitin	264604

735	11607959 (1469, 1470)	Novel Protein sim. GBank gi|401582|sp|P27432|YICE_ECOLI - HYPOTHETICAL 48.9 KD PROTEIN IN GLTS-SELC INTERGENIC REGION			264594
736	10879734 (1471, 1472)	Novel Protein sim. GBank gi|400831|sp|P31135|POTH_ECOLI - PUTRESCINE TRANSPORT SYSTEM PERMEASE PROTEIN POTH	Contains protein domain (PF00528) - Binding-protein-dependent transport systems inner membrane component	transport	264636
737	78945340 (1473, 1474)		Contains protein domain (PF00615) - Regulator of G protein signaling domain	UNCLASSIFIED	265020
738	17895353 (1475, 1476)				265008
739	79833670 (1477, 1478)	Novel Protein sim. GBank gi|2506867|sp|P33225|TORA_ECOLI - TRIMETHYLAMINE-N-OXIDE REDUCTASE PRECURSOR (TMAO REDUCTASE) (TRIMETHYLAMINE OXIDASE)		oxidase	264910
740	19881557 (1479, 1480)				264907, 264764, 264634, 264637
741	79827273 (1481, 1482)	Novel Protein sim. GBank gi|3261828|emb|CAB10925| - (Z98260) mrp [Mycobacterium tuberculosis]	Contains protein domain (PF01883) - Domain of unknown function	UNCLASSIFIED	264689, 35696286, 264510, 264908, 18108362
742	82393795 (1483, 1484)	Novel Protein sim. GBank gi|3877494|emb|CAA88472.1| - (Z48583) ATP binding protein with similarity to the CDC48/PAS1/SEC18 family; cDNA EST EMBL:D65037 comes from this gene; cDNA EST EMBL:D68340 comes from this gene; cDNA EST EMBL:D65048 comes from this gene; cDNA EST EMBL:D6845...		UNCLASSIFIED	29331822, 264910, 264762
743	82300051 (1485, 1486)	Novel Protein sim. GBank gi|127420|sp|P19888|MTBA_BACAR - MODIFICATION METHYLASE BANI (CYTOSINE-SPECIFIC METHYLTRANSFERASE BANI) (M.BANI)	Contains protein domain (PF00145) - C-5 cytosine-specific DNA methylase	UNCLASSIFIED	264488, 264259, 264508, 264905, 264906, 264907, 264908, 264909, 264510, 264511, 264512, 265008, 265009, 264910, 264591, 264596, 264759, 265010, 265011, 18108351, 264763, 264288, 264766, 264768, 264693, 18108370, 264629, 18108372, 264630, 264631, 264634, 264558, 18108385, 264482, 264564, 264567
744	80230421 (1487, 1488)				18108397, 264511, 264690, 264628, 264638, 264692, 264639, 264766
745	9841963 (1489, 1490)	Novel Protein sim. GBank gi|78921|pir||S04846 - UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate--D-alanyl-D-alanine ligase (EC 6.3.2.15) precursor - Escherichia coli		glycoprotein	264906
746	11073229 (1491, 1492)	Novel Protein sim. GBank gi|3386354 (AF074705) - pyochelin synthetase [Pseudomonas aeruginosa]		synthase	264600
747	94322044 (1493, 1494)	Novel Protein sim. GBank gi|2887411|dbj|BAA24848| - (AB007878) KIAA0418 [Homo sapiens]	Contains protein domain (PF00018) - SH3 domain	oxidase	66714117, 264905, 264509, 264906, 264907, 264908, 264909, 264511, 264910, 265011, 264681, 264288, 264766, 264687, 264768, 264769, 21906768, 35695917, 264691, 264693, 264628, 264634, 264635, 264639, 56182323, 83373044
748	11617923 (1495, 1496)				264690